Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?

Sanne Maria Nabuurs

Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?

Sanne Maria Nabuurs

Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?

Sanne M. Nabuurs

Some books are to be tasted, others to be swallowed, and some few to be chewed and digested; that is, some books are to be read only in parts; others to be read but not curiously; and some few to be read wholly, and with diligence and attention.

Sir Francis Bacon

Table of contents

1

General Introduction

Sanne M. Nabuurs and Carlo P. M. van Mierlo

The history of protein folding

In 1839 the Dutch chemist Gerhardus Johannes Mulder carried out elemental analysis of "animal substances", like fibrin, albumin and gelatin, and discovered they all shared a similar chemical composition [1]. Jöns Jakob Berzelius advised him in a letter written in French to call the substance "protéine" after the Greek word πρωτειος, meaning "primary", since these substances are the primary compounds of animals and also need to be primary compounds of plants, which are eaten by animals [2].

Until 1902 it was unclear how proteins were build-up from their building blocks, the amino acids. It was known that hydrolysis could break proteins down, but the nature of the linkage between the amino acids was obscure. Already in 1901 Emil Fischer produced the dipeptide glycineglycine (and thereby also established the peptide nomenclature) [3]. Finally, on the 22nd of September 1902 Franz Hofmeister and Emil Fischer both gave a lecture at the annual meeting of the prestigious Gesellschaft der deutscher Naturforscher und Ärtze in Karlsbad [4]. Both scientists proposed that in proteins the constituent amino acids are joined to each other by the condensation of the amino group of one amino acid with the carboxyl group of another amino acid to form amide bonds in a linear structure [5, 6]. It was still unclear whether proteins were molecules or colloidal aggregates. If proteins were molecules they were much bigger than any molecule known. Only with the discovery of the ultracentrifuge in 1924 this issue was solved, proteins like hemoglobin had a large molecular weight and all particles were of the same size [7]. In the 30's the concept of proteins being macromolecules had become firmly established.

The first protein denaturation experiments were performed in the beginning of the 20th century. In 1900 Spiro observed that coagulation of protein solutions by raising the temperature was delayed by adding urea, furthermore, heat coagulated proteins could be dissolved again by urea [8]. In 1902 Ramsden placed a dead frog in a saturated urea solution and observed that it "becomes translucent and falls to pieces in a few hours" [9].

The elucidation of the existence of the peptide bond could not solve the question why proteins are compact and globular. The discipline of protein folding was established in the 30's upon appearance of a lucid paper in the Chinese Journal of Physiology by Dr. Hsien Wu, in which he writes the following still up-to-date statement [10, 11]:

> 'Evidence is adducted in support of the hypothesis that the molecule of natural, soluble protein is not a flexible open chain of polypeptide but has a compact structure. The force of attraction between the polar groups in a single molecule of protein holds them together in an orderly way, just as the force of attraction between different molecules holds many molecules together in a crystal. In denaturation or coagulation the compact and orderly structure is disorganized. If denaturation occurs in acid or alkali or in urea solution, the individual molecules are disrupted but they remain separate. In coagulation they interpenetrate and are entangled. The facts known about denaturation and coagulation in diverse ways are explained and correlated by the theory.'

Hydrogen bonds became well known later in the 1930's and in 1936 Mirsky and Pauling wrote "On the Structure of Native, Denatured, and Coagulated Proteins" [12] without having knowledge about the paper of Wu. They state:

> *'The denatured protein molecule we consider to be absent of a uniquely defined configuration.' 'We consider the native protein to be held in its definite configuration by side-chain hydrogen bonds'*

In 1934 Mirsky, among others, demonstrated the reversibility of the denaturation of some proteins [13]. In the same year Bernal and Crowfoot discovered that pepsin crystals gave excellent diffraction patterns [14], however, there was no way of analyzing the data in detail with the techniques known at the time, but they could still observe that the structure was compact and closely packed. More than 20 years later the low-resolution structure of myoglobin was published [15] and refined to 2 Å [16] in 1960. This structure proved the statement of Wu in 1931 that weak non-covalent bonds hold the 3D structure of a protein together. Hydrophobic effects in proteins were known to exist in the 30s [17], however, the importance of hydrophobicity in protein folding became widely recognized after a now famous paper of Kauzmann in 1959 [18].

The knowledge that proteins have defined three-dimensional structures didn't solve the question how such three-dimensional structures were formed on basis of their amino acid sequences. In 1958 it was known that "DNA makes RNA makes protein", and that the newly synthesized product of RNA is an inactive, unfolded polypeptide chain. How does this chain fold and become active? In 1961 Anfinsen demonstrated the reversibility of denaturation of ribonuclease A, a protein containing four disulphide bonds [19]. He showed that the amino acid sequence contains the information to form the sole correct set of disulphide bonds out of 105 possibilities. Later in 1973 when he had won the Nobel Prize in Chemistry he proposed a thermodynamic hypothesis: the newly synthesized polypeptide chain folds driven by the free energy of conformation that is gained in going to the stable, native structure [20].

In the same decade the question arose whether the folding of proteins was two-state, or accomplished through involvement of intermediate states. Brandts proposed the two-state model for protein unfolding, where the native and unfolded conformations are in equilibrium, based on the data for chymotrypsinogen [21]. This model is still valid today for interpreting equilibrium folding reactions for many small proteins. The idea that proteins can fold without the formation of intermediate states triggered Cyrus Levinthal in 1968 to calculate the time it would take a protein to search randomly all possible backbone conformations. He estimated this search would take longer than the lifetime of the universe [22, 23]. Levinthal's paradox, as it became known, suggested that folding intermediates and folding pathways must exist.

The protein-folding problem

In the 21st century the question how proteins fold is still not completely solved. How does the amino acid sequence of a protein determine its final three-dimensional structure? This question is known as the protein-folding problem, and can be separated into different components: Can the three-dimensional structure of a protein be predicted from its amino acid sequence? What are the mechanisms and pathways by which proteins fold from an unfolded to their folded state? Why does

a protein with a specific amino acid sequence fold to one particular three-dimensional structure instead to some other structure?

These questions all refer to the difference between unfolded (U) and native (N) proteins, and the conformations a protein may form on route from U to N. The thermodynamic hypothesis of Afinsen [20] states that the native state of a protein in its physiological milieu is the one with the lowest free energy. The fast and direct folding of proteins was explained by Levinthal trough the use of the concept of protein folding pathways, i.e., a protein folds along a defined pathway from unfolded to native protein via well-defined partially structured intermediate states [22].

In the 90s a new view of protein folding arose: the concept of the multidimensional energy landscape or "folding funnel" [24, 25]. To make this idea clear Dill and Chan used the metaphor of a mountainside, where the flow of water down this mountainside is less like the flow trough a simple gulley (folding pathway) but more like the trickle of water down mountainsides of complex shapes [25]. The vertical axis of a folding funnel represents the free energy of a conformation (Fig. 1.1), while the multiple horizontal axes represent different degrees of freedom of a protein chain, such as the dihedral angles. Structurally similar conformations will be close in space on the landscape. In some experimentally or in silico determined energy landscapes the horizontal axes display different properties of a protein chain, such as the radius of gyration or root mean square deviation [26]. Changing conditions such as temperature and solvent will stretch or shrink the landscape.

Figure 1.1 [25] illustrates different folding concepts. The situation in which an unfolded protein randomly searches to find the native state, like a golf-ball in a flat golf-course, is shown in Figure 1.1a. In this situation it will take a long time for the ball to find the hole and drop down into it. The folding pathway Levinthal suggested is illustrated in Figure 1.1b, where an unfolded protein positioned at **A** goes through a tunnel to the native state **N**, this process is thus more directed than the random search shown in Figure 1.1a. According to Levinthal: 'A pathway of folding means that there exists a well-defined sequence of events which follow one another.' [22]

However, the folding of a protein does not start from one specific conformation, **A**. The denatured or unfolded state is the collection of all possible conformations except the native state **N**. In Figure 1.1c an idealized smooth protein folding energy landscape is shown, in which all unfolded conformations reach the native state with the same rate, since all slopes are equal, but not necessarily via the same sequence of conformations.
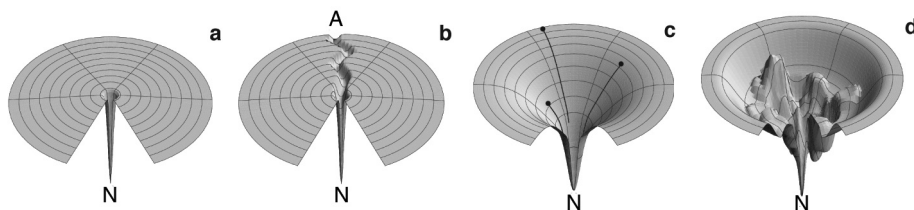
Figure 1.1 Energy landscapes, also called folding funnels. (a) The Levinthal 'golf-course'. **N** is the native conformation. The chain searches for **N** randomly. (b) The 'pathway' solution to the random coil search problem. A pathway is assumed to lead from an unfolded conformation **A** to the native conformation **N**, conformational searching is more directed and folding is faster than for random searching. (c) An idealized funnel landscape, as the chain forms increasing numbers of intrachain contacts, and lowers its internal free energy, its conformational freedom is also reduced. (d) A rugged energy landscape with kinetic traps, energy barriers, and some narrow paths to the native conformation. All figures are taken from [25].

Figure 1.1d illustrates the folding of a non-ideal protein, i.e., the protein has an energy landscape with multiple intermediates and transition states. Turning to the metaphor of a mountain [25], proteins can be depicted as skiers. Under denaturing conditions an ensemble of skiers is distributed everywhere on a mountainside (the presence of denaturant will reshape the funnel to a more flat energy landscape, since the free energy of the native state is no longer the lowest). A stopped-flow jump that lowers the denaturant concentration and induces protein folding recreates a folding funnel and the skiers will proceed downhill all following a different route. Valleys they encounter are kinetic traps of 'misfolded' intermediates and the uphill slopes will cause the skiers to slow down. The tops are the transition states, which have to be passed on route from the intermediate states to the native state. In this manner the folding energy landscape guides the assembly of three-dimensional structures through Brownian motions. The free energy of a given structure on the three-dimensional landscape decreases as the protein gets closer to the native ensemble at the bottom of the funnel.

Quite some effort has been made to construct three-dimensional folding landscapes for real proteins based on experimental data. The landscape of α-lactalbumin has been constructed from NMR data obtained at different concentrations urea between 0 and 10 M, with the radius of gyration and the root mean square deviation from the native state as horizontal axes. A realistic multidimensional folding landscape consisting of all possible chain conformations of a protein is still theoretical and creating this landscape will be a major step in solving the protein-folding problem.

The unfolded state (U)
The completely unfolded protein is a random coil. All rotation angles of the backbone and side-chains can change independent of all other rotation angles and all possible conformations have comparable free energies, except when atoms come too close in proximity of one another. Steric repulsions between atoms in the covalent structure limit local flexibility and as no spatial overlap between atoms can exist, a fraction of all possible conformations is excluded. Despite these restrictions, the total number of conformations possible for a random coil protein is astronomical, for a small protein with 100 amino acids and an average of 8 conformations per residue there would be up to 10^{89} conformations [27].

Unfolded proteins in strong denaturants like 6 M guanidine hydrochloride (GuHCl) or 8 M Urea have the hydrodynamic properties of a random coil, as was shown in 1966 by Tanford and co-workers [28]. They showed that the intrinsic viscosity ($[\eta]$) of 16 denatured proteins of different chain lengths (n) fall on a straight line plot of $\log[\eta]$ versus log n with a slope of 0.666, which is close to the value of 0.6 that Flory predicted in his random coil theory [29, 30]. Recent NMR [31] and SAXS [32] studies of denatured proteins experimentally confirm the value of 0.6, predicted by Flory. In a truly random coil the energetics of interactions between different parts of the polypeptide have to be equally balanced by the interactions the polypeptide has with the solvent. Since a protein is constituted of 20 different amino acids, it is unlikely that these amino acids all have the same interaction with the solvent and with one another. It is thus impossible for a protein to behave as an "ideal" random coil. Indeed, residual structure has been found in several denatured proteins, even under harsh conditions as 6 M GuHCl or 8 M urea, while the hydrodynamic radii of these denatured proteins are still indistinguishable from those expected for the corresponding random coils [33, 34, 35]. How can this random coil behavior arise? It was recently demonstrated by using Monte Carlo simulations that ensembles of a polypeptide chain that is largely comprised of pre-organized non-random segments have end-to-end distances and mean radii of gyration that agree well with

random coil expectations [36]. Thus, simulations show that an unfolded protein can have residual structure while still showing random coil dimensions.

Structural and dynamical characterization of unfolded proteins is of importance for understanding protein folding since the unfolded state defines the starting point for folding reactions. The denatured state of a protein in water is described by Baldwin as "an embryonic folding intermediate" [37]. Residual structure in unfolded proteins in presence or in absence of denaturants can be studied by various techniques. The overall shape and size of an unfolded protein is revealed by SAXS (small-angle x-ray scattering) [38], structural and dynamical features of an unfolded protein are obtained by NMR spectroscopy [39, 40], while FRET (Förster resonance energy transfer) and FCS (fluorescence correlation spectroscopy) can be used to study single unfolded protein molecules [41].

The native state (N)

Native, folded proteins have complex conformations, and their structures are revealed in great detail by X-ray crystallography [15] and NMR spectroscopy [40, 41]. At this moment, more than 50.000 protein structures have been deposited to the Protein Data Bank [42]. While most structurally characterized proteins have a defined shape and structure, large part of our (and of other organisms) genome translates to proteins that are not well structured in solution [43]. These proteins are called "intrinsically disordered" or "natively unfolded" and are involved in signal transduction or transcription, for instance. Handling of these proteins is difficult since production and purification leads to severe experimental difficulties. Characterization of these proteins is as difficult as characterization of folding intermediates and unfolded proteins, since little or no fixed structure is present.

Two-state folding

Originally, several small single-domain proteins were found to fold according to a simple two-state folding mechanism. In case of these "two-state folders" both the native and unfolded state can co-exist at equilibrium with one another. Upon adding denaturant these two-state folders typically display a highly cooperative (un)folding transition linking N and U (Fig. 1.2a). The folding reaction $U \leftrightarrows N$ is characterized by a single folding rate constant (k_N) and by a single unfolding rate constant (k_U) and the ratio of these rate constants equals the equilibrium constant K_{U-N}:
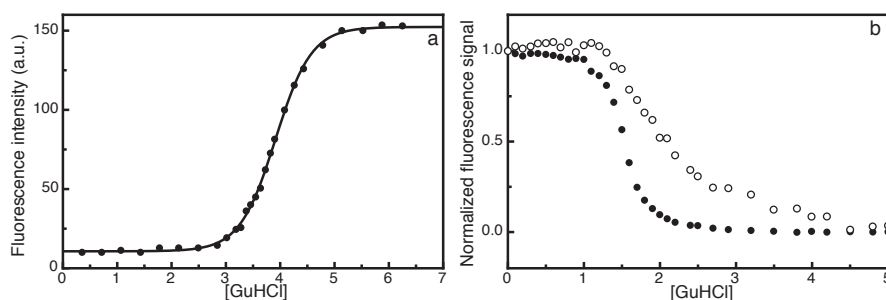

Figure 1.2 GuHCl-induced equilibrium unfolding data of CI2 and *Azotobacter vinelandii* apoflavodoxin.
(a) Equilibrium unfolding curve of CI2 measured by intrinsic fluorescence at 356 nm (data taken from [44])
(b) Equilibrium unfolding curves of apoflavodoxin, as measured by intrinsic fluorescence at 340 nm • and by far-UV CD at 222 nm o (data taken from [45]).

$$K_{U\text{-}N} = k_U/k_N \qquad\qquad [1.1]$$

In analogy with chemical reactions, the two folding states are separated by an energy barrier, the so-called transition state. The transition state is a conformation along the reaction path between U and N that has the highest free energy (Fig. 1.3). The standard free energy of denaturation ΔG^o is described by:

$$\Delta G^o = -RT \ln K_{U\text{-}N} \qquad\qquad [1.2]$$

To analyze urea- or GuHCl-induced unfolding curves, like the one shown in Figure 1.2a, the linear extrapolation method (LEM) can be used. This method enables the determination of ΔG^o which is found to vary linearly with denaturant concentration according to [46]:

$$\Delta G = \Delta G^o - m[D] \qquad\qquad [1.3]$$

where ΔG^o is an estimate of the conformational stability of the native protein at 0 M denaturant, and m is a measure of the dependence of ΔG on denaturant concentration.

Equation 1.4 is used to obtain a nonlinear least-square fit of experimentally detected unfolding data, the equation contains the identity of Equation 1.3 and takes into account the linear dependence of the pre- and post-unfolding data on denaturant concentration:

$$y = \frac{(a_N + b_N[D]) + (a_U + b_U[D]) \cdot e^{\frac{-(\Delta G^o - m[D])}{RT}}}{1 + e^{\frac{-(\Delta G^o - m[D])}{RT}}} \qquad\qquad [1.4]$$

with a_N and a_U the intercepts, b_N and b_U the slopes of the pre- and post-transition baselines, and ΔG^o and m as defined by Equation 1.3.

The kinetics of folding and refolding can be deciphered by measuring folding and unfolding rates using a range of denaturant concentrations. These rates are extrapolated to zero molar denaturant to estimate folding and unfolding rates in absence of denaturant. The plot of the logarithm of the
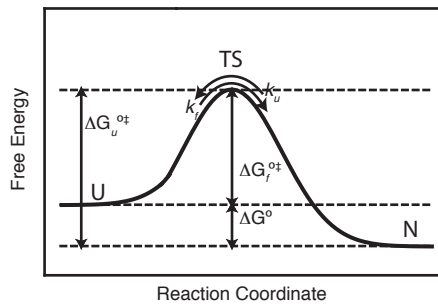

Figure 1.3 Free-energy profile for protein folding and unfolding. Shown are the relationship between the free energies of activation for folding ($\Delta G^{o\ddagger}_f$), and unfolding ($\Delta G^{o\ddagger}_u$) and the equilibrium free energy (ΔG^o), using conventional transition-state theory.

observed rates versus denaturant concentration resembles a *V*; hence, this plot is commonly referred to as a chevron plot (Fig 1.4a).

New and improved experimental methods have reduced the number of small proteins that exhibit true two-state folding behavior in kinetic and equilibrium (un)folding experiments to nearly zero. Ultra-rapid mixing [47] led to the detection of short-lived kinetic intermediates and techniques such as FRET and NMR spectroscopy revealed rare species during the folding of several small proteins [48, 49].
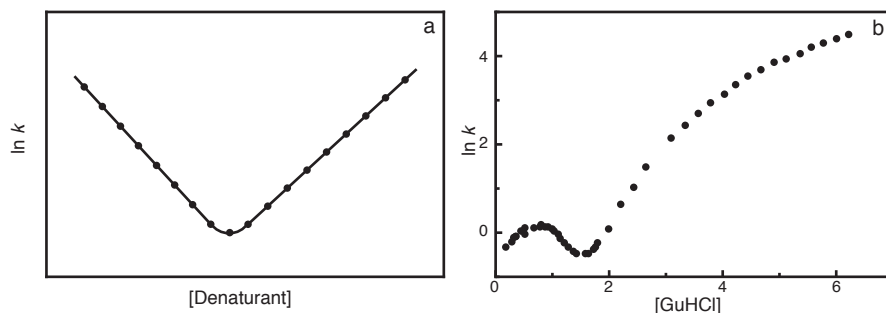

Figure 1.4 Chevron plot shows the denaturant dependence of the observed rate constant for (un)folding (ln *k*). (a) Classical two-state folding behavior, during which intermediates are not populated. (b) Chevron plot obtained for *A. vinelandii* apoflavodoxin, the roll-over at low concentration denaturant reveals the presence of an off-pathway intermediate, the non-linear unfolding limb at high concentration denaturant reveals the presence of a high-energetic on-pathway intermediate [45].

Intermediate or "molten globule" states

Levinthal predicted the existence of a folding pathway on which folding intermediates reside. In the 70s the search for such intermediate states began and in 1973 while studying the denaturation of Bovine Carbonic Anhydrase B, Wong and Tanford discovered "separable sequential conformational transitions" [50], meaning that they detected the presence of a folding intermediate. In 1976 a folding intermediate of α-lactalbumin was discovered by Kuwajima and co-workers [51]. Ptitsyn stated that the folding intermediates found were the same species as the partly folded structures formed under mildly destabilizing conditions, such as acidic pH [52]. Ohgushi and Wada gave the name "molten globule" to these partially folded structures in 1983; "globule" meaning compact, "molten" meaning having no fixed tertiary structure [53]. A molten globule is characterized by (i) the presence of a substantial amount of secondary structure, (ii) the virtual absence of tertiary structure associated with tight packing of side chains, (iii) the relative compact size of the molecule, a radius only 10-30% larger than that of the native state, (iv) a loosely packed hydrophobic core and a hydrophobic surface exposed to solvent, and (v) fluctuations of the structure on a time-scale longer than nanoseconds [54]. Molten globule states have been found for many proteins under mildly denaturing conditions; some are studied extensively, such as the molten globules of α-lactalbumin [55-57], apomyoglobin [58-60] and of cytochrome c [53, 61, 62].

Intermediate states populated at equilibrium can be detected by comparing equilibrium-unfolding curves obtained with different techniques, such as tryptophan fluorescence and far-UV CD spectroscopy. Tryptophan fluorescence reports about the tertiary microenvironment surrounding tryptophans in a protein, since water exposed tryptophans have lower intrinsic

fluorescence compared to tryptophans shielded in a folded structure. Far-UV CD spectroscopy detects primarily α-helical content in a protein. If the unfolding curves obtained by both techniques do not coincide, a folding intermediate is present as shown in Figure 1.2b.

Stopped-flow mixing experiments can reveal involvement of kinetic intermediates, which are not populated to great extent at equilibrium, but are transiently formed during folding or unfolding of proteins. By making use of spectroscopic techniques, like tryptophan fluorescence, the exponential decay or increase of a spectroscopic signal is measured. Two-state folders will give rise to single exponentially decaying or increasing curves, whereas proteins that fold via one or more intermediate states give rise to multi-exponential curves.

The structural characterization of folding intermediates is experimentally challenging, since most intermediates are hardly populated relative to the population of native and unfolded states. However, folding intermediates can be trapped by appropriately changing solvent conditions [63] or by making use of site-directed mutagenesis [64-66].

Off-pathway intermediates

Off-pathway intermediates are non-productive species in the folding energy landscape of a protein. They need to unfold before productive folding to the native state can occur. These intermediate species have been detected during kinetic folding of for instance apoflavodoxin [45], tryptophan synthase [67] and UMP/CMP kinase [68]. Kinetic off-pathway species are revealed by examining the chevron plots of folding and unfolding. The presence of an off-pathway intermediate gives rise to a curvature in the folding limb of the chevron plot (Fig. 1.4b). Upon increasing denaturant concentration instead of lowering the folding rate, this rate actually increases, which is characteristic for an unfolding step (i.e., the unfolding of an off-pathway intermediate).

In the perspective of a multidimensional folding energy landscape, the term "off-pathway" is conceptually strange. An off-pathway species still lies somewhere on the mountainside in the energy landscape. However, the conformation of these species may be totally different from the native state, and consequently high energy mountaintops on which more unfolded conformations reside will separate these species from the native state.

Protein folding in vivo

Folding of proteins *in vivo* is in many aspects different from folding *in vitro*. *In vivo* folding starts co-translationally [69], while *in vitro* (re)folding involves the whole protein molecule. The environment a folding protein experiences in the cell is completely different from the one in the test tube. A cell is extremely crowded with macromolecules, the protein concentration can be as high as 300 mg/ml [70]. Due to the high concentration of proteins close to the ribosomal exit tunnel, newly formed proteins are susceptible to aggregation. To avoid the latter, folding of newly synthesized polypeptides to their native conformation is guided by the sequential action of multiple molecular chaperones [71, 72]. The synthesis of a typical protein in a cell can take up to several minutes [73], whereas protein folding occurs on a much faster timescale. Failure of correct folding in the cell is undesirable, not only because of loss of function of a protein, but more importantly because of the toxicity of protein aggregates formed by misfolded proteins [74].

NMR spectroscopy as a tool to study protein folding

One of the best tools to gain detailed structural and dynamical information about folding states residing within a folding energy landscape is NMR spectroscopy. Non-native states of proteins do not adopt unique three-dimensional structures as folded proteins do, but instead are ensembles of rapidly fluctuating conformations. Despite their conformational heterogeneity, unfolded and intermediate folding states can be studied by different NMR techniques. In 1992 the first unfolded protein was characterized by NMR: the urea unfolded N-terminal domain of 434-repressor. By measuring NOE restraints, residual structure was found in this unfolded protein [75].

Over the years several NMR techniques have been used to study unfolded and intermediate folding states of proteins. The dynamical behavior of unfolded proteins and folding intermediates usually differs a lot. Unfolded proteins exhibit very rapid dynamics and conformational exchange on the (sub)nanosecond timescale leading to conformational averaging, and the observation of very sharp cross peaks in the corresponding NMR spectra. In contrast, the ensemble of conformations that forms a folding intermediate or "molten globule" state are characterized by exchange between conformations on the micro- to milli-second timescale, which leads to extensive broadening of the corresponding NMR resonances. In the following paragraphs an overview is given of the NMR techniques most often used to study unfolded and intermediate folding states of proteins.

Resonance assignments

Any detailed investigation of conformation and dynamics of proteins at atomic resolution by NMR spectroscopy requires the assignment of the resonances of amino acids involved. ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence) spectra of unfolded proteins show typical poor dispersion in the H^N direction. In contrast, the dispersion in the ^{15}N direction is quite good. A doubly labeled sample (^{13}C-^{15}N) of an unfolded protein allows one to use the well-dispersed ^{13}C' resonances for sequential resonance assignments. Triple resonance experiments are used to assign backbone nuclei by correlating nuclei of a specific residue i with nuclei of the neighbouring amino acid i-1. Some of the most used three-dimensional triple resonance experiments and their correlations are illustrated in Figure 1.5.
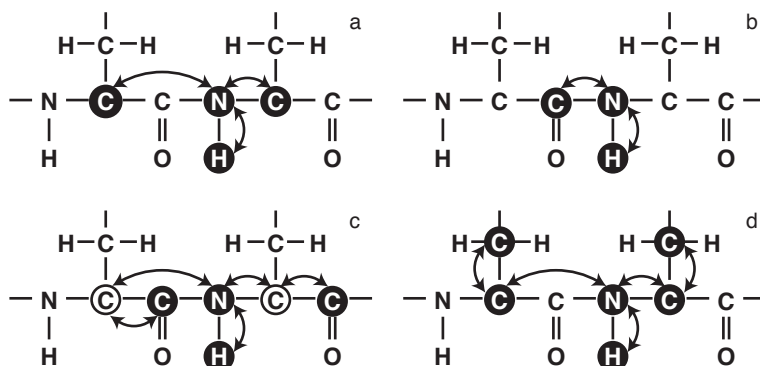
Figure 1.5 Triple resonance experiments used for sequential resonance assignment. (a) HNCA, correlations observed: ^{1}H^N_i-^{15}N$_i$-^{13}C$^\alpha_i$, ^{1}H^N_i-^{15}N$_i$-^{13}C$^\alpha_{i-1}$ (b) HNCO, correlations observed: ^{1}H^N_i-^{15}N$_i$-^{13}C'$_{i-1}$ (c) HN(CA)CO, correlations observed: ^{1}H^N_i-^{15}N$_i$-^{13}C'$_i$, ^{1}H^N_i-^{15}N$_i$-^{13}C'$_{i-1}$ (d) HNCACB, correlations observed: ^{13}C$^\beta_i$/^{13}C$^\alpha_i$-^{15}N$_i$-^{1}H^N_i, ^{13}C$^\beta_{i-1}$/^{13}C$^\alpha_{i-1}$-^{15}N$_i$-^{1}H^N_i

Chemical shifts

The chemical shift (δ) depends on the environment of the observed nucleus and is a sensitive predictor of secondary structure [76, 77]. Backbone resonances of unfolded proteins have chemical shifts close to random coil chemical shifts. These random coil chemical shifts have been experimentally determined by NMR measurements using small peptides in the form of GGXA [78], GGXGG [79, 80] or GGYXGG [81, 82] (with Y being either a proline or glycine and X any of the 20 amino acids) in either aqueous or denaturing solutions. Chemical shift deviations from these random coil chemical shifts are often called secondary shifts.

Chemical shifts are sensitive to secondary structure formation. Uninterrupted stretches of positive or negative secondary shifts indicate helix- or strand-formation, respectively. Both $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts are shifted downfield when the corresponding atoms are in α-helical structures and upfield, when the corresponding atoms are located in β-sheets. $^{13}C^{\beta}$, ^{15}N, H^{α} and H^{N} resonances experience upfield shifts in α-helices and downfield shifts in β-sheets.

The most sensitive predictor of α-helices is the $^{13}C^{\alpha}$ secondary shift. In folded proteins secondary shifts of 2-3 ppm are observed for $^{13}C^{\alpha}$ located in an α-helix. In unfolded proteins the secondary shifts of transient helices can be much smaller, the secondary shifts are usually less than 1 ppm and often these shifts are only a few tenths of a ppm. Stretches of positive or negative secondary shifts, even if small in value, indicate the presence of residual structure in unfolded proteins. In unfolded proteins $^{13}C'$ is well dispersed, originating from sequence-dependent effects. While $^{13}C'$ chemical shift deviations from random coil values are as sensitive to structure formation as the $^{13}C^{\alpha}$ chemical shifts are, the sequence dependency of $^{13}C'$ makes the use of random coil values difficult. Fortunately, intrinsic referencing, i.e., using random coil values of the same polypeptide, solves this problem [40].

Secondary shifts have been used for the structural characterization of several unfolded proteins and folding intermediates. Residual structure has been found in the unfolded states of for instance apomyoglobin [34, 83] and lysozyme [33].

Nuclear Overhauser Effects

NOEs form the basis for the classical three-dimensional structure determination of proteins by NMR spectroscopy. They provide long-range constraints on the global topology, and short-range constraints on local structure, including secondary structure [84]. In unfolded proteins it is hard to measure long-range constraints, but short-range NOEs have been observed for several unfolded proteins [85-87].

Structures of folding intermediates can be solved by NOEs when their resonances are not severely broadened due to conformational exchange. The structures of folding intermediates of En-HD [64] and Rd-apocytochrome b_{562} [88] were solved using both long- and short-range NOEs.

Heteronuclear relaxation

Backbone dynamics of proteins can be characterized at the residue level by NMR spectroscopy. The most commonly used practice for studying polypeptide chain dynamics of unfolded proteins and folding intermediates involves measurement of ^{1}H-^{15}N T_{1}, T_{2} and heteronuclear NOE relaxation data. Analysis of these data is in general similar to the analysis of folded proteins. However, classical model-free analysis [89, 90] of the data is usually not valid, since the assumption of a single overall rotational correlation time is invalid for partly or fully unfolded proteins [91].

The ^{1}H-^{15}N NOE is sensitive to motions on the picosecond to nanosecond timescale. Folded proteins have NOE ratios of 0.8-0.9. In case of partly folded and unfolded proteins, the NOE

ratio can range from negative values up to 0.7-0.8 for the structured parts of molten globules [92]. R_2 relaxation rates are sensitive to nanosecond motions and to exchange processes on the milli- to microsecond timescale. R_1 relaxation rates are often rather uniform in partly folded and unfolded proteins. The method of reduced spectral density mapping [93, 94], in which no assumption of overall tumbling is required, can be applied to analyze the heteronuclear relaxation data. However, this analysis does not always provide additional insights into the backbone motions of unfolded and partly folded proteins [92].

Relaxation dispersion experiments can be used to detect intermediate folding states that are populated at low levels (i.e., at about 0.5% or higher). Both chemical shift differences between two states as well as (un)folding kinetics can be extracted from the experimental data [95]. A folding intermediate of Fyn SH3 domain was detected using this technique. As chemical shifts of the intermediate were determined by relaxation dispersion, a structural model of the intermediate state could also be proposed [96]. Relaxation dispersion profiles of ACBP at different concentrations denaturant were acquired to gain insight into the folding and unfolding kinetics of this protein [97]. The authors constructed a chevron plot of folding and unfolding rates and revealed a folding intermediate of ACBP.

Paramagnetic relaxation enhancements

Long-range interactions in proteins can be probed by using covalently attached nitroxide spin labels. These paramagnetic spin labels cause broadening of nuclear spins within a radius of about 15 Å [98, 99] and were originally used to determine distances in folded proteins. The method of paramagnetic relaxation enhancements (PRE) was first used on (partly) unfolded proteins by studying a denatured fragment of staphylococcal nuclease [100] and since then frequently used to study unfolded and partly folded proteins, such as α-synuclein [101, 102], ACBP [103, 104] and apomyoglobin [105, 106].

In PRE experiments paramagnetic spin labels such as MTSL, CMTLS and PROXYL are attached at specific sides of a protein. The labels react with the thiol group of a single cysteine residue, which can be present in wild-type protein or can be introduced by site-directed mutagenisis. ^{1}H-^{15}N HSQC spectra of the protein with the label in the paramagnetic (oxidized spin label) and diamagnetic (reduced spin label) state are recorded. Often ascorbic acid is used to reduce the spin label. The differences in line widths of resonances or differences in cross peak intensities give an estimate of the average distance between the spin label and any given amide.

H/D exchange

Backbone amide protons of proteins dissolved in D_2O will exchange with deuterons, and upon complete exchange they are no longer not observable in ^{1}H-^{15}N HSQC spectra. The intrinsic kinetics of this exchange process depends on pH, temperature and the neighbouring residues of a particular backbone amide investigated [107]. If an amide is involved in a hydrogen bond the rate of exchange will be decreased compared to the predicted intrinsic exchange rate. Hydrogen exchange rates of unfolded proteins will be almost as high as the intrinsic exchange rates [108]. Partly folded proteins contain secondary structure and amide protons in these structured regions will be protected against exchange. This phenomenon was first demonstrated for the acid-induced molten globule of α-lactalbumin [109].

Rare folding species or Partly Unfolded Forms (PUFs) of a protein can be detected by H/D experiments using folded proteins [110]. For example, H/D experiments of apoflavodoxin in the

presence of small amounts of denaturant indentified the presence of four PUFs. Two of these PUFs are partly folded forms of the off-pathway folding intermediate of apoflavodoxin [111].

To probe hydrogen exchange rates of species with rapidly exchanging amide protons, quenched-flow hydrogen exchange experiments have been developed [112, 113, 114]. In this methodology, the intermediate state of a protein is trapped in D_2O, for instance by changing the solvent, and after time t the protein is folded to its native state (quenched) in which no (or very slow) hydrogen exchange occurs. Subsequently, ^{1}H-^{15}N HSQC spectra of the native protein are recorded to determine the corresponding exchange rates. By choosing the different times t properly, an exponential decay of a backbone amide proton intensity reveals the hydrogen exchange rate of the amide proton in the intermediate folding state. This method was used to examine the adsorbed state of BLA, which has conformational properties similar to BLA molecules in the acid-induced molten globule state [113].

Short-lived folding intermediates that cannot be trapped at equilibrium can also be detected by quenched-flow hydrogen exchange. These folding intermediates are pulse-labeled during the folding process and can be detected by analyzing NMR spectra of the final folded protein [115]. Unfortunately, amide protons that are positioned at the periphery of native proteins exchange relatively rapidly and thus it is not possible to probe the corresponding amide protons in folding intermediates. To overcome this problem a new method has been developed, as demonstrated in the study folding intermediates of apomyoglobin. Instead of measuring NMR spectra of native protein after pulse labeling, the protein is rapidly lyophilized and dissolved in DMSO, where no additional hydrogen exchange can occur. This methodology resulted in a more complete picture of a folding intermediate of apomyoglobin, since now exchange rates of additional parts of the folding intermediate could be determined [114].

Flavodoxins and flavodoxin folding

Flavodoxins are small monomeric flavoproteins, which function as low-potential one-electron carriers in bacteria and contain a non-covalently bound FMN cofactor [116]. These proteins adopt the α-β parallel topology, in which a central parallel β-sheet is surrounded by α-helices (Fig. 1.6). This topology is often called the flavodoxin-like fold, although flavodoxins share the fold with proteins that are sequently and functionally unrelated, like catalases, cutinases, esterases and chemotactic proteins [117].

In 1964 a flavoprotein containing flavin mononucleotide (FMN) was isolated from the obligate aerobe bacterium *Azotobacter vinelandii* [118]. Initially it was difficult to demonstrate its function. Therefore the protein has been known by several names: the Shetna protein [119], after its discoverer, Azotobacter free-radical protein [120]

Figure 1.6 Cartoon drawing of flavodoxin from *A. vinelandii* (pdb ID 1YOB), FMN cofactor is shown in dark gray.

and Azotoflavin [121]. Based on structural similarities, the protein was believed to be a flavodoxin, but it took several years before it was found to be a flavodoxin on basis of an activity assay [122]. In contrast to other flavodoxins known at the time, *A. vinelandii* flavodoxin's main function is not to replace ferrodoxin under low iron conditions, but it is instead involved in transport of electrons to

nitrogenase, a nitrogen-fixating enzyme [121]. Later, it was demonstrated that *A. vinelandii* produces three flavodoxins, with flavodoxin II being the most abundant under nitrogen fixating conditions [123].

The folding behavior of *A. vinelandii* apo- and holoflavodoxin II has been studied extensively during the past years. Both denaturant-induced equilibrium and kinetic (un)folding of apoflavodoxin have been characterized in detail using GuHCl as denaturant [46, 110, 121-123]. The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to flavodoxin except for considerable dynamics of residues in the flavin-binding region [124, 125]. In presence of FMN, binding of the FMN cofactor to native apoflavodoxin is the last step in flavodoxin folding.

Apoflavodoxin kinetic folding involves an energy landscape with two folding intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [45]. Intermediate I_{on} is an obligatory species on the productive route from unfolded to native protein and is highly unstable and is therefore not observed during denaturant-induced equilibrium unfolding. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [45]. Apoflavodoxin equilibrium unfolding observed by fluorescence emission, fluorescence anisotropy and far-UV CD spectroscopy is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin [45]. Intermediate I_{off} populates significantly in the concentration range of 1 to 3 M GuHCl (Fig. 1.7).

The off-pathway species is molten globule-like: it is compact, it lacks the characteristic structure of native apoflavodoxin, its three tryptophans are solvent exposed as opposed to the situation in native protein, it has at least 65% of the α-helical content of native apoflavodoxin, and it is highly dynamic with severely broadened NMR resonances due to exchange between different conformers on the micro- to millisecond timescale [45]. The formation of such an off-pathway species, which needs to unfold to allow folding to proceed to the native state, is typical for apoflavodoxin's topology [117]. Severe aggregation of this species occurs at elevated protein concentration [126] and also due to macromolecular crowding [127].
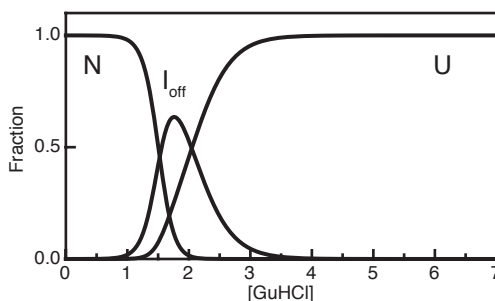

Figure 1.7 Fractions of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as function of GuHCl concentration

Outline of this thesis

The conformational properties of the unfolded state of *Azotobacter vinelandii* flavodoxin and its off-pathway, molten globule-like, folding intermediate are presented in this thesis. *A. vinelandii* flavodoxin is a representative of α-β parallel proteins, which are widely prevalent in nature. The kinetic and equilibrium folding of *A. vinelandii* flavodoxin have been described in detail. Like other α-β parallel proteins, *A. vinelandii* flavodoxin populates an off-pathway intermediate. As a result of the characterization of the conformational properties of unfolded apoflavodoxin and apoflavodoxin's molten globule, the question: "Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?" is answered in this thesis.

Chapter 2 of the thesis describes characterization of GuHCl-unfolded apoflavodoxin at the residue-level using heteronuclear NMR spectroscopy. It turns out that in 6.0 M denaturant, the protein behaves as a random coil. In contrast, at 3.4 M denaturant, ^{13}C secondary shifts and ^{1}H-^{15}N relaxation rates report four transiently ordered regions in unfolded apoflavodoxin. These regions have restricted flexibility on the (sub)nanosecond timescale. Secondary shifts show that three of these regions form α-helices, which are populated about 10% of the time, as confirmed by far-UV CD data. One region of unfolded apoflavodoxin adopts non-native structure. Of the α-helices observed, two are present in native apoflavodoxin as well. A substantial part of the third helix becomes β-strand while forming native protein. Chemical shift changes due to amino acid residue replacement show that the latter α-helix has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin. Remarkably, these ordered segments dock non-natively, which causes strong competition with on-pathway folding. Thus, rather than directing productive folding, chapter 2 shows that conformational pre-organization in the unfolded state of an α-β parallel type protein promotes off-pathway species formation.

The off-pathway intermediate of flavodoxin can be populated at equilibrium using GuHCl as denaturant. This allows detection of molten globule formation at the residue level by using heteronuclear NMR spectroscopy. **Chapter 3** presents changes in chemical shifts of backbone amides, as well as disappearance of resonances, of unfolded apoflavodoxin upon decreasing denaturant concentration. Analysis of the data shows that structure formation within virtually all parts of the unfolded protein precedes folding to the molten globule state. This folding transition is non-cooperative and involves a series of distinct transitions. Four structured elements in unfolded apoflavodoxin transiently interact and subsequently form the ordered core of the molten globule. Although hydrophobic, tryptophan side chains are not involved in the latter process. In addition, chapter 3 shows that this ordered core is gradually extended upon decreasing denaturant concentration, but part of apoflavodoxin's molten globule remains random coil in the denaturant range investigated.

To further investigate long-range interactions in unfolded apoflavodoxin, which are involved during formation of the off-pathway intermediate, use is made of site-directed spin labeling in **chapter 4** of this thesis. For the purpose of spin labeling, glutamine at position 48 of apoflavodoxin is replaced by cysteine. Residue 48 resides in the non-native a-helix of unfolded apoflavodoxin. Introduction of this cysteine enables attachment of different nitroxide spin labels to the protein, such as MTSL and CMTSL. It turns out that the stability of native apoflavodoxin against unfolding decreases by this amino acid replacement and a further decrease in stability is caused by subsequent attachment of a nitroxide spin label. Due to increased hydrophobic interactions, replacement of Gln48 by Cys48 also decreases flexibility of particular regions in unfolded apoflavodoxin in

3.4 M GuHCl. Interpretation of the data obtained is further complicated due to non-specific interactions between nitroxide spin labels and hydrophobic patches within unfolded apoflavodoxin. Consequently, chapter 4 shows that one has to be careful in the use of spin labels for the study of unfolded proteins. The spin label data show that non-native contacts exist between transiently ordered structured elements in unfolded apoflavodoxin.

Full population of the molten globule state of apoflavodoxin, which is made possible by replacing Phe44 with Tyr44, is reported in **chapter 5** of this thesis. It is demonstrated that this replacement leads to severe destabilization of native apoflavodoxin and only marginally affects the molten globule. As a result, F44Y apoflavodoxin forms the molten globule at low salt concentration, facilitating characterization of its conformation. It is shown that the molten globule is helical and contains no β-sheet. Its conformation thus differs drastically with respect to the one of native apoflavodoxin, which contains a parallel β-sheet surrounded by α-helices. The presence of this remarkably non-native species shows that a single polypeptide sequence can form considerably different folds. Chapter 5 proofs that upon changing conditions, conformational switching between two distinct folds can occur.

Full population of the molten globule state of F44Y apoflavodoxin enables use of H/D exchange for the characterization at the residue level by NMR spectroscopy of apoflavodoxin's molten globule folding intermediate. In **chapter 6**, interrupted H/D exchange is used to detect the stable core of apoflavodoxin's molten globule in absence of denaturant. It is shown that Leu110 to Val125 have the highest protection factors against H/D exchange and form the single stable core of apoflavodoxin's molten globule in absence of denaturant. In contrast, chapter 3 shows that in presence of GuHCl, four regions of apoflavodoxin form the folding core of the molten globule. Clearly, the helical region Leu110 to Val125 is better buried from solvent than the other three ordered regions of apoflavodoxin's molten globule. It is tempting to speculate that this burial is due to residues Leu110 to Val125 being surrounded by the remaining three ordered regions of the off-pathway folding intermediate of apoflavodoxin.

References

1. Mulder, G. J. (1839). Ueber die Zusammensetzung einiger thierischen Substanzen. *Journal für Praktische Chemie* **16**, 129-152.
2. Hartley, H. (1951). Origin of the Word 'Protein'. *Nature* **168**, 244.
3. Fischer, E. (1901). Über einige Derivate des Glykocolls. *Berichte der Deutschen Chemischen Gesellschaft* **34**, 2868-2871.
4. Fruton, J. S. (1985). Contrasts in scientific style. Emil Fischer and Franz Hofmeister: their research groups and their theory of protein structure. *Proc Am Philos Soc* **129**, 313-370.
5. Hofmeister, F. (1902). Über Bau und Gruppierung der Eiweisskörper. *Reviews of Physiology, Biochemistry and Pharmacology* **1**, 759-802.
6. Fischer, E. (1902). Über die Hydrolyse der Proteinstoffe. *Chemiker-Zeitung* **26**, 939-940.
7. Svedberg, T. & Fahraeus, R. (1926). A new method for the determination of the molecular weight of the proteins. *J Am Chem Soc* **48**, 430-438.
8. Spiro, K. (1900). Ueber die Beeinflussung der Eiweisscoagulation durch stickstoffhaltige Substanzen. *Zeitschrift fuer physiologische Chemie* **30**, 182-199.
9. Ramsden, W. (1902). Some new properties of urea. *Journal of Physiology* **28**, xxiii-xxvi.
10. Wu, H. (1931). Studies on denaturation of proteins. XIII. A theory of denaturation. *Chinese Journal of Physiology* **5**, 321-344.
11. Edsall, J. T. (1995). Wu, Hsien and the First Theory of Protein Denaturation (1931). *Adv Protein Chem* **46**, 1-26.
12. Mirsky, A. E. & Pauling, L. (1936). On the Structure of Native, Denatured, and Coagulated Proteins. *Proc Natl Acad Sci U S A* **22**, 439-447.
13. Mirsky, A. E. & Anson, M. L. (1934). The equilibrium between native and denatured hemoglobin in salicylate solutions and the theoretical consequences of the equilibrium between native and denatured protein. *J Gen Physiol* **17**, 393-398.
14. Bernal, J. D. & Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature* **133**, 794.
15. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666.
16. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **185**, 422-427.
17. Edsall, J. T. (1992). Memories of early days in protein science, 1926-1940. *Protein Sci* **1**, 1526-1530.
18. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**, 1-63.
19. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309-1314.
20. Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science* **181**, 223-230.
21. Brandts, J. F. (1964). The Thermodynamics of Protein Denaturation. I. The Denaturation of Chymotrypsinogen. *J Am Chem Soc* **86**, 4291-4301.
22. Levinthal, C. (1968). Are there pathways for protein folding? *J Chim Phys* **65**, 44-45.
23. Levinthal, C. (1969). *Mössbauer Spectroscopy in Biological Systems*, Allerton House, Monticello, Illinois.
24. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.
25. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19.
26. Vendruscolo, M., Paci, E., Karplus, M. & Dobson, C. M. (2003). Structures and Relative Free Energies of Partially Folded States of Proteins. *Proc Natl Acad Sci U S A* **100**, 14817-14821.
27. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv Protein Chem* **33**, 167-241.
28. Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J Biol Chem* **241**, 1921-1923.
29. Flory, P. J. (1953). *Principles of Polymer Chemistry*, Cornell University Press.
30. Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*, Wiley, New York.
31. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-16431.
32. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S. & Plaxco, K. W. (2004).

Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci U S A* **101**, 12491-12496.

33.   Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J. & Dobson, C. M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**, 8977-8991.

34.   Schwarzinger, S., Wright, P. E. & Dyson, H. J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry* **41**, 12681-12686.

35.   Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **36**, 2390-2402.

36.   Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci U S A* **101**, 12497-12502.

37.   Baldwin, R. L. (2002). A new perspective on unfolded proteins. *Adv Protein Chem* **62**, 361-367.

38.   Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S., Forman-Kay, J. D. & Kay, L. E. (2002). Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J Mol Biol* **316**, 101-112.

39.   Kumar, A., Srivastava, S., Mishra, R. K., Mittal, R. & Hosur, R. V. (2006). Local structural preferences and dynamics restrictions in the urea-denatured state of SUMO-1: NMR characterization. *Biophys J* **90**, 2498-2509.

40.   Modig, K., Jurgensen, V. W., Lindorff-Larsen, K., Fieber, W., Bohr, H. G. & Poulsen, F. M. (2007). Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett* **581**, 4965-4971.

41.   Schuler, B. & Eaton, W. A. (2008). Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* **18**, 16-26.

42.   Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucl Acids Res* **28**, 235-242.

43.   Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, 739-756.

44.   Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435.

45.   Bollen, Y. J., Sanchéz, I. E. & van Mierlo, C. P. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475-10489.

46.   Greene, R. F., Jr. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**, 5388-5393.

47.   Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-1861.

48.   Kay, L. E. (2005). NMR studies of protein structure and dynamics. *J Magn Reson* **173**, 193-207.

49.   Schuler, B. (2005). Single-molecule fluorescence spectroscopy of protein folding. *ChemPhysChem* **6**, 1206-1220.

50.   Wong, K. P. & Tanford, C. (1973). Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. *J Biol Chem* **248**, 8518-8523.

51.   Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976). Three-state denaturation of α-lactalbumin by guanidine hydrochloride. *J Mol Biol* **106**, 359-373.

52.   Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981). α-lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett* **136**, 311-315.

53.   Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* **164**, 21-24.

54.   Maki, K., Kamagata, K. & Kuwajima, K. (2005). Equilibrium and Kinetically Observed Molten Globule States. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 856-883. 5 vols. WILEY-VCH Verlag GmbH & Co. KGaA, Wienheim.

55.   Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S. & Dobson, C. M. (1999). Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat Struct Biol* **6**, 948-952.

56.   Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol* **4**, 630-634.

57.   Quezada, C. M., Schulman, B. A., Froggatt, J. J., Dobson, C. M. & Redfield, C. (2004). Local and global cooperativity in the human alpha-lactalbumin molten globule. *J Mol Biol* **338**, 149-158.

58. Eliezer, D., Yao, J., Dyson, H. J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol* **5**, 148-155.

59. Goto, Y., Takahashi, N. & Fink, A. L. (1990). Mechanism of acid-induced folding of proteins. *Biochemistry* **29**, 3480-3488.

60. Eliezer, D., Chung, J., Dyson, H. J. & Wright, P. E. (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* **39**, 2894-2901.

61. Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. & Roder, H. (1996). Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry* **35**, 5538-5549.

62. Colon, W. & Roder, H. (1996). Kinetic intermediates in the formation of the cytochrome c molten globule. *Nat Struct Biol* **3**, 1019-1025.

63. Eliezer, D., Jennings, P. A., Dyson, H. J. & Wright, P. E. (1997). Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR. *FEBS Lett* **417**, 92-96.

64. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053-1056.

65. Whittaker, S. B., Spence, G. R., Gunter Grossmann, J., Radford, S. E. & Moore, G. R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* **366**, 1001-1015.

66. Schwarzinger, S., Mohana-Borges, R., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2008). Structural characterization of partially folded intermediates of apomyoglobin H64F. *Protein Sci* **17**, 313-321.

67. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X. & Matthews, C. R. (2007). A Tightly Packed Hydrophobic Cluster Directs the Formation of an Off-pathway Sub-millisecond Folding Intermediate in the α Subunit of Tryptophan Synthase, a TIM Barrel Protein. *J Mol Biol* **366**, 1624-1638.

68. Lorenz, T. & Reinstein, J. (2008). The Influence of Proline Isomerization and Off-Pathway Intermediates on the Folding Mechanism of Eukaryotic UMP/CMP Kinase. *J Mol Biol* **381**, 443-455.

69. Chen, W., Helenius, J., Braakman, I. & Helenius, A. (1995). Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc Natl Acad Sci U S A* **92**, 6229-6233.

70. Zimmerman, S. B. & Trach, S. O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* **222**, 599-620.

71. Maier, T., Ferbitz, L., Deuerling, E. & Ban, N. (2005). A cradle for new proteins: trigger factor at the ribosome. *Curr Opin Struct Biol* **15**, 204-212.

72. Young, J. C., Agashe, V. R., Siegers, K. & Hartl, F. U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* **5**, 781-791.

73. Thulasiraman, V., Yang, C. F. & Frydman, J. (1999). In vivo newly translated polypeptides are sequestered in a protected folding environment. *EMBO J* **18**, 85-95.

74. Saibil, H. R. (2008). Chaperone machines in action. *Curr Opin Struct Biol* **18**, 35-42.

75. Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.

76. Wang, Y. & Jardetzky, O. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **11**, 852-861.

77. Wishart, D. S. & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol* **239**, 363-92.

78. Bundi, A. & Wüthrich, K. (1979). 1H NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* **18**, 285-297.

79. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E. & Dyson, H. J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR* **18**, 43-8.

80. Plaxco, K. W., Morton, C. J., Grimshaw, S. B., Jones, J. A., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. (1997). The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. *J Biomol NMR* **10**, 221-230.

81. Merutka, G., Dyson, H. J. & Wright, P. E. (1995). Random Coil H-1 Chemical-Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series GGXGG. *J Biomol NMR* **5**, 14-24.

82. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). H-1, C-13 and N-15 Random Coil NMR Chemical-Shifts of the Common Amino-Acids. 1. Investigations of Nearest-Neighbor Effects. *J Biomol NMR* **5**, 67-81.

83. Yao, J., Chung, J., Eliezer, D., Wright, P. E. & Dyson, H. J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40**, 3561-3571.
84. Wüthrich, K. (1986). NMR of proteins and nucleic acids. Wiley, Amsterdam, New York.
85. Penkett, C. J., Redfield, C., Jones, J. A., Dodd, I., Hubbard, J., Smith, R. A., Smith, L. J. & Dobson, C. M. (1998). Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from Staphylococcus aureus. *Biochemistry* **37**, 17054-17067.
86. Tafer, H., Hiller, S., Hilty, C., Fernandez, C. & Wuthrich, K. (2004). Nonrandom Structure in the Urea-Unfolded Escherichia coli Outer Membrane Protein X (OmpX). *Biochemistry* **43**, 860-869.
87. Le Duff, C. S., Whittaker, S. B., Radford, S. E. & Moore, G. R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: implications for the early stages of protein folding. *J Mol Biol* **364**, 824-835.
88. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci U S A* **102**, 5026-5031.
89. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546-4559.
90. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* **104**, 4559-4570.
91. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem Rev* **104**, 3607-3622.
92. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-132.
93. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* **6**, 153-162.
94. Peng, J. W. & Wagner, G. (1991). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *J Magn Reson* **98**, 308-332.
95. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* **123**, 11341-11352.
96. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-590.
97. Teilum, K., Poulsen, F. M. & Akke, M. (2006). The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein. *Proc Natl Acad Sci U S A* **103**, 6877-6782.
98. Kosen, P. A., Scheek, R. M., Naderi, H., Basus, V. J., Manogaran, S., Schmidt, P. G., Oppenheimer, N. J. & Kuntz, I. D. (1986). Two-dimensional 1H NMR of three spin-labeled derivatives of bovine pancreatic trypsin inhibitor. *Biochemistry* **25**, 2356-2364.
99. Schmidt, P. G. & Kuntz, I. D. (1984). Distance measurements in spin-labeled lysozyme. *Biochemistry* **23**, 4261-4266.
100. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J Mol Biol* **268**, 170-184.
101. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. (2005). Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **127**, 476-477.
102. Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M. & Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Natl Acad Sci U S A* **102**, 1430-1435.
103. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M. & Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J Mol Biol* **347**, 1053-1062.
104. Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M. & Vendruscolo, M. (2004). Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J Am Chem Soc* **126**, 3291-3299.
105. Lietzow, M. A., Jamin, M., Jane Dyson, H. J. & Wright, P. E. (2002). Mapping long-range contacts in a highly unfolded protein. *J Mol Biol* **322**, 655-662.

106. Felitsky, D. J., Lietzow, M. A., Dyson, H. J. & Wright, P. E. (2008). Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci U S A* **105**, 6278-6283.

107. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.

108. Buck, M., Radford, S. E. & Dobson, C. M. (1994). Amide hydrogen exchange in a highly denatured state. Hen egg-white lysozyme in urea. *J Mol Biol* **237**, 247-254.

109. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* **28**, 7-13.

110. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-197.

111. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* **103**, 4095-4100.

112. Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694-699.

113. Engel, M. F., Visser, A. J. & van Mierlo, C. P. (2004). Conformation and orientation of a protein folding intermediate trapped by adsorption. *Proc Natl Acad Sci U S A* **101**, 11316-11321.

114. Nishimura, C., Dyson, H. J. & Wright, P. E. (2005). Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc Natl Acad Sci U S A* **102**, 4765-4770.

115. Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* **335**, 700-704.

116. Mayhew, S. G. & Tollin, G. (1992). General properties of flavodoxins. In *Chemistry and biochemistry of flavoenzymes* (Müller, F., ed.), pp. 389-426. CRC Press, Boca Raton, FL.

117. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

118. Shethna, Y. I., Wilson, P. W., Hansen, R. E. & Beinert, H. (1964). Identification by isotopic substitution of the EPR signal at g = 1.94 in a non-heme iron protein from azotobacter *Proc Natl Acad Sci U S A* **52**, 1263-1271.

119. Tollin, G. & Edmondson, D. E. (1971). Flavoprotein chemistry. II. Chemical and physical characterization of the Shethna flavoprotein and apoprotein and kinetics and thermodynamics of flavine analog binding to the apoprotein. *Biochemistry* **10**, 124-132.

120. Hinkson, J. W. & Bülen, W. A. (1967). A Free Radical Flavoprotein from Azotobacter. Isolation, crystallization, and properties. *J Biol Chem* **242**, 3345-3351.

121. Yoch, D. C., Benemann, J. R., Valentine, R. C. & Arnon, D. I. (1969). The electron transport system in nitrogen fixation by Azotobacter. II. Isolation and function of a new type of ferredoxin. *Proc Natl Acad Sci U S A* **64**, 1404-1410.

122. van Lin, B. & Bothe, H. (1972). Flavodoxin from *Azotobacter vinelandii*. *Arch Microbiol* **82**, 155-172.

123. Klugkist, J., Voorberg, J., Haaker, H. & Veeger, C. (1986). Characterization of three different flavodoxins from *Azotobacter vinelandii*. *Eur J Biochem* **155**, 33-40.

124. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of *Azotobacter vinelandii* holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

125. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

126. van Mierlo, C. P. M., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

127. Engel, R., Westphal, A. H., Huberts, D. H., Nabuurs, S. M., Lindhoud, S., Visser, A. J. & van Mierlo, C. P. M. (2008). Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem* **283**, 27383-27394.

2

Extensive formation of off-pathway species during folding of an α-β parallel protein is due to docking of (non)native structure elements in unfolded molecules

Abstract

Detailed information about unfolded states is required to understand how proteins fold. Knowledge about folding intermediates formed subsequently is essential to get grip on pathological aggregation phenomena. During folding of apoflavodoxin, which adopts the widely prevalent α-β parallel topology, most molecules fold via an off-pathway folding intermediate with helical properties. To better understand why this species is formed, guanidine hydrochloride-unfolded apoflavodoxin is characterized at the residue-level using heteronuclear NMR spectroscopy. In 6.0 M denaturant, the protein behaves as a random coil. In contrast, at 3.4 M denaturant, secondary shifts and ^{1}H-^{15}N relaxation rates report four transiently ordered regions in unfolded apoflavodoxin. These regions have restricted flexibility on the (sub)nanosecond timescale. Secondary shifts show that three of these regions form α-helices, which are populated about 10% of the time, as confirmed by far-UV CD data. One region of unfolded apoflavodoxin adopts non-native structure. Of the α-helices observed, two are present in native apoflavodoxin as well. A substantial part of the third helix becomes β-strand while forming native protein. Chemical shift changes due to amino acid residue replacement show that the latter α-helix has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin. Remarkably, these ordered segments dock non-natively, which causes strong competition with on-pathway folding. Thus, rather than directing productive folding, conformational pre-organization in the unfolded state of an α-β parallel type protein promotes off-pathway species formation.

Introduction

Folding of proteins to conformations with proper biological activities is of vital importance to all living organisms. Current knowledge about kinetic and energetic aspects of protein folding largely stems from in vitro studies and simulations, see e.g. [1-3]. To describe folding, the concept of a multi-dimensional free energy landscape or folding funnel arose [4]. In this model, unfolded protein molecules descend along a funnel describing their free energy until they reach the state that has the lowest free energy, which is the native state. In the energy landscape model, unfolded protein molecules can fold to the native state by following different routes.

We use the expression 'unfolded state' to refer to the ensemble of unfolded protein molecules at the top of the folding funnel. This state is thermodynamically separated from other folding states and corresponds to a given macroscopically observable quantity. The mean dimensions of a chemically denatured protein are those of a random coil ensemble [5]. However, it has become apparent that several unfolded proteins contain residual structure [6, 7]. Simulations show that unfolded proteins with non-random segments can indeed exhibit random coil statistics [8]. Despite its importance, experimental information at the atomic-level about structural properties of unfolded proteins is relatively sparse, see e.g. [6, 9-13]. This is due to conformational heterogeneity of the unfolded state and dynamic interconversion between different conformers of unfolded molecules.

Upon folding, proteins can encounter rough folding energy landscapes that allow population of partially folded species, which may be on- or off-pathway to the native state. To gain insight into events that occur early during folding detailed information about unfolded states is crucial. Residual structure in the unfolded state is important as it most likely facilitates formation of folding intermediates. Knowledge about unfolded states is also relevant because some unfolded and partially folded proteins possess biological function [14] or are involved in aggregation phenomena that play a pathological role in several human diseases [15].

Here, we investigate the guanidine hydrochloride (GuHCl)-induced unfolded state of a 179-residue flavodoxin from *Azotobacter vinelandii*. Flavodoxins consist of a single structural domain that adopts the α-β parallel topology. This topology is characterized by a parallel β-sheet flanked by α-helices (Fig. 2.1a). Unraveling flavodoxin folding contributes to understanding how the many proteins with an α-β parallel topology fold.

Native apoflavodoxin (i.e., flavodoxin without non-covalently bound FMN cofactor) is structurally largely identical to flavodoxin [16, 17] and contains no disulphide bonds. Both the GuHCl-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin have been

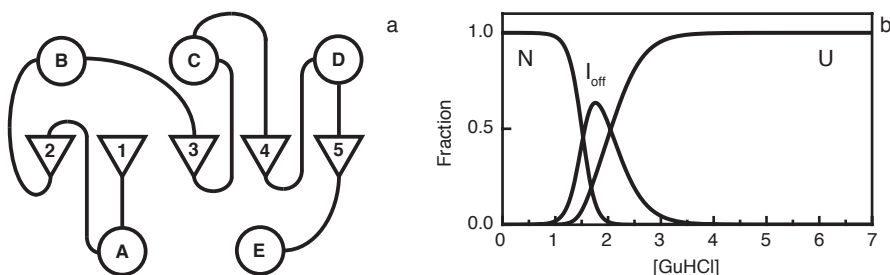*Figure 2.1* (a) Schematic drawing of the α-β parallel topology. α-Helices are represented as circles and β-strands as triangles. (b) Fractions of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as function of GuHCl concentration [18].

characterized [18-20]. Apoflavodoxin kinetic folding involves an energy landscape with two folding intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [18]. Intermediate I_{on} lies on the productive route from unfolded to native protein and is highly unstable and is thus not observed during denaturant-induced equilibrium unfolding. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and acts as a trap [18]. Apoflavodoxin equilibrium unfolding is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin [18]. Intermediate I_{off} populates significantly in the concentration range of 1 to 3 M GuHCl (Fig. 2.1b). The off-pathway species is compact, helical and molten globule-like and has extremely broadened NMR resonances [18, 21]. At elevated protein concentration, severe aggregation of this species occurs [21].

Population of an off-pathway intermediate seems typical for folding of proteins with an α-β parallel topology [22]. Residual structure in the corresponding unfolded states probably facilitates formation of this species. Here, detailed information about unfolded apoflavodoxin is revealed by NMR spectroscopy. Our observations show that ordered segments exist in unfolded apoflavodoxin and that these segments dock non-natively, leading to off-pathway species formation. Apparently, especially proteins that contain domains with an α-β parallel topology are susceptible to off-pathway intermediate formation.

Materials and Methods

Sample preparation. The single cysteine at position 69 in wild-type *A. vinelandii* (strain ATCC 478) flavodoxin II was replaced by an alanine (Cys69Ala) to avoid covalent dimerization of apoflavodoxin. Uniformly ^{15}N and ^{13}C-^{15}N labeled flavodoxin was obtained from transformed *E. coli* cells grown on ^{15}N and ^{13}C-^{15}N labeled algae medium (Silantes), and purified as described [44].

Unfolded apoflavodoxin was obtained by denaturing flavodoxin in 6 M GuHCl. Subsequently, FMN was removed via gel filtration at 7 M GuHCl. Two NMR samples of 1.5 mM unfolded ^{13}C-^{15}N labeled Cys69Ala apoflavodoxin in 3.4 M or 6.0 M GuHCl, respectively, were prepared in 10% D_2O. Samples of 0.5 mM Cys69Ala and Gln48Cys/Cys69Ala ^{15}N labeled apoflavodoxin, both unfolded in 3.4 M GuHCl, were prepared. Gln48Cys/Cys69Ala apoflavodoxin was generated using site-directed mutagenesis. 2,2-Dimethyl-2-silapentane-5-sulfonic acid was present as internal chemical shift reference. DTT was present in the sample of Gln48Cys/Cys69Ala apoflavodoxin. Refractometry was used to verify GuHCl concentration [45]. The buffer used was 100 mM potassium pyrophosphate, pH 6.0.

Far-UV CD. CD measurements were performed on a Jasco J715 spectropolarimeter. Spectra ranging from 210 to 260 nm were recorded of 10 μM apoflavodoxin unfolded in buffer containing 3.4 M or 6.0 M GuHCl, respectively, and appropriate blank spectra were subtracted. Samples were measured in a 1 mm quartz cuvette (Starna) at 25 ºC. A total of 80 scans, each comprising 501 data points, were recorded per sample.

NMR spectroscopy. Spectra were recorded on a Bruker Avance 700 MHz machine and on a Bruker DMX 500 MHz machine. Sample temperature was 25 ºC. A series of heteronuclear NMR experiments (Tables 2.1 and 2.2) was acquired to assign backbone resonances of unfolded apoflavodoxin in 3.4 M or 6.0 M GuHCl, respectively (for pulse sequences see [46]).

	max. evolution times (ms)[a]			Number of complex points		
Experiment	t_1, max	t_2, max	t_3, max	n_1	n_2	n_3
¹H-¹⁵N HSQC	139 (N)	244 (H)	-	512	4096	-
HNCA	12.2 (C)	15.5 (N)	292 (H)	120	110	4096
HNCO	35 (C)	13.5 (N)	292 (H)	100	96	4096
HN(CA)CO	35 (C)	15.5 (N)	292 (H)	100	110	4096
HNCACB	4.7 (C)	15.5 (N)	292 (H)	100	110	4096
CBCACONH	2.4 (C)	13.5 (N)	292 (H)	100	96	4096
T_2 (CPMG)	36 (N)	122 (H)	-	256	2048	-

[a](N), (C) and (H) refer to the time axis for ¹⁵N, ¹³C, and ¹H, respectively.
Table 2.1 Acquisition parameters for NMR experiments with apoflavodoxin unfolded in 6.0 M GuHCl.

	max. evolution times (ms)[a]			Number of complex points		
Experiment	t_1, max	t_2, max	t_3, max	n_1	n_2	n_3
¹H-¹⁵N HSQC	139 (N)	244 (H)	-	512	4096	-
HNCO	35 (C)	13.5 (N)	292 (H)	100	96	4096
HN(CA)CO	35 (C)	15.5 (N)	292 (H)	100	110	4096
HNCACB	4.7 (C)	15.5 (N)	292 (H)	100	110	4096
CBCACONH	2.4 (C)	13.5 (N)	292 (H)	100	96	4096
T_1	69 (N)	122 (H)	-	256	2048	-
T_2 (CPMG)	36 (N)	122 (H)	-	256	2048	-
¹H-¹⁵N NOE	73 (N)	122 (H)	-	256	2048	-
$T_{1\rho}$	69 (N)	122 (H)	-	256	2048	-

[a](N), (C) and (H) refer to the time axis for ¹⁵N, ¹³C, and ¹H, respectively.
Table 2.2 Acquisition parameters for NMR experiments with apoflavodoxin unfolded in 3.4 M GuHCl.

Uniformly ¹⁵N labeled apoflavodoxin unfolded in 3.4 M GuHCl was used to determine ¹⁵N longitudinal (T_1), transverse (T_2) relaxation rates and ¹H-¹⁵N NOE values [46, 47]. T_1 relaxation times were obtained from a series of spectra using the following delays in the corresponding pulse sequence: 100, 200 (2x), 300, 500, 752 (2x), 1000, 1500 and 2000 ms, respectively. T_2 relaxation times were measured by using the CPMG sequence. A series of spectra was acquired on both a 500 MHz and a 700 MHz spectrometer using the following CPMG mixing times: 8, 16 (2x), 32, 64, 100, 152 (2x), 200, 252, 300 and 400 ms, respectively. A relaxation delay of 3 s was used during all measurements. The NOE measurements were carried out using a 3 s saturation period after a 6 s recycle delay. Experiments with and without presaturation were alternately done and in total 12 spectra were acquired.

Data analysis. All spectra were processed with NMRPipe [48] and analyzed using NMRViewJ [49]. The relaxation rates were determined by least squares fitting a single-exponentially decaying function to the cross peak intensity as a function of the delay in the corresponding pulse sequence. Duplicate experiments were done to estimate the standard deviations of the cross peak intensities. NOE values were calculated from the peak intensity ratios of spectra with and without proton saturation.

The intrinsic flexibility of a residue is simulated using the radius of gyration R_g of the side chain involved [50]. The intrinsic correlation time τ for each amino acid is assumed proportional to R_g^3, following Stoke's Law. The decreased conformational freedom associated with proline is modeled by increasing its effective side chain R_g to 2.0 Å [50]. Relaxation rate R_2 at residue i, i.e., R_{2i}, is calculated using Equation 2.1 [23], in which the effect of neighboring amino acids on the backbone motions of each residue is assumed to decay exponentially [24]:

$$R_{2i} = k \sum_{j=1}^{N} \tau_j \cdot e^{\frac{-|i-j|}{\lambda_j}}$$

[2.1]

where k is an empirical scaling constant, N is the total number of residues, τ_j is the intrinsic correlation time of residue j, and λ_j is the persistence length for segmental motion of the polypeptide chain. The persistent length λ_j is set to 7 residues for all amino acid residues [24], except for glycine and alanine for which the value of λ_j is set to 2 to account for the high flexibility of these residues [23].

Values for the average area buried upon folding (AABUF) [33] were calculated using ProtScale [51]. A window size of 9 and a relative weight of the window edges compared to the window center of 10% were used.
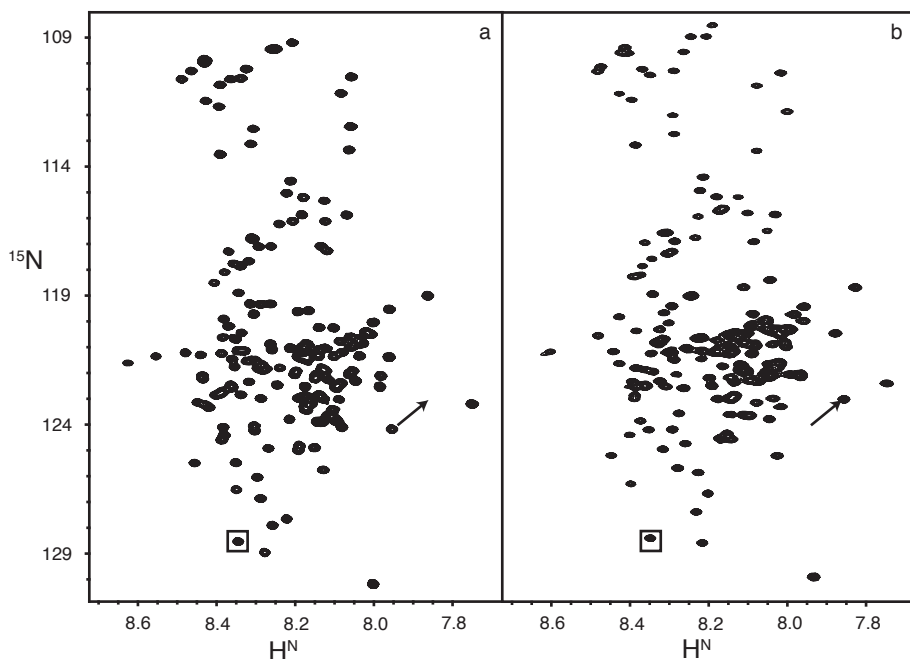
Figure 2.2 Gradient-enhanced 700 MHz ¹H-¹⁵N HSQC spectra of 1.5 mM ¹³C-¹⁵N-labeled apoflavodoxin in 6.0 M (a) or 3.4 M (b) GuHCl. The box in (a) and (b) highlights an example of a cross peak (Ala18) that does not shift upon lowering the GuHCl concentration from 6.0 M to 3.4 M GuHCl. The large shift of the cross peak of Ala169 upon lowering the denaturant concentration is shown by arrows.
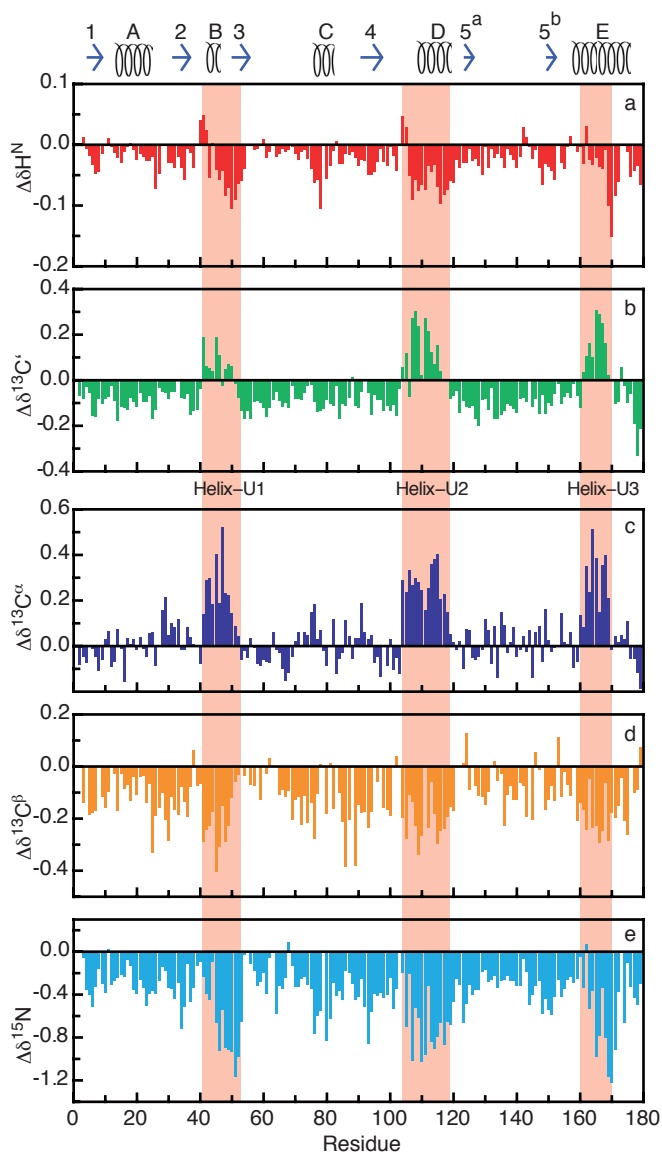
Figure 2.5 Secondary shifts of apoflavodoxin in 3.4 M GuHCl calculated by intrinsic referencing. Secondary shifts are shown for H^N (a), $^{13}C'$ (b), $^{13}C^\alpha$ (c), $^{13}C^\beta$ (d), ^{15}N (e). The regions of apoflavodoxin that form α-helical structures in 3.4 M GuHCl (i.e., Helix-U1, Helix-U2 and Helix-U3) are highlighted. α-Helices and β-strands in native flavodoxin are shown at the top.

Results

Upon lowering denaturant concentration, the amount of residual structure in a chemically unfolded protein is expected to increase and become better detectable. The lowest concentration GuHCl at which unfolded apoflavodoxin is the only observable species is 3.4 M (Fig. 2.1b). Consequently, we chose this concentration to reveal potential conformational pre-organization in unfolded apoflavodoxin. To maximize random coil-like behavior of apoflavodoxin, 6.0 M GuHCl is used.

Resonance assignments of unfolded apoflavodoxin. The ^{1}H-^{15}N HSQC spectra of unfolded apoflavodoxin in 3.4 M and in 6.0 M GuHCl are fingerprints of unfolded apoflavodoxin. Both spectra have limited ^{1}H chemical shift dispersion, which is typical for unfolded proteins (Fig. 2.2). Assignments of all corresponding H^N, ^{15}N, ^{13}C', ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances are obtained using heteronuclear NMR experiments (See Materials and Methods). The assignments (BioMagResBank, accession number 15474 and Supporting Information, Tables S2.1 and S2.2) are facilitated due to the relatively large dispersion of backbone ^{15}N and ^{13}C' chemical shifts.
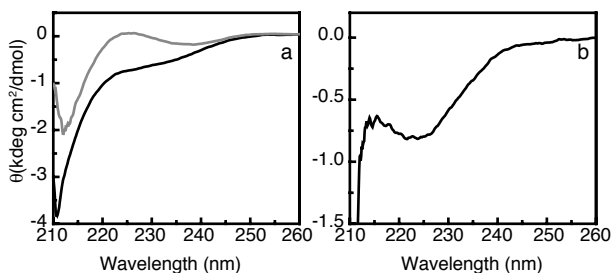

Figure 2.3 Far-UV CD data of unfolded apoflavodoxin. (a) The spectrum of apoflavodoxin in 6.0 M GuHCl (gray line) is typical for random coil protein. The spectrum of apoflavodoxin in 3.4 M GuHCl is shown in black. (b) Difference CD spectrum obtained by subtracting the gray spectrum from the black one in (a).

Apoflavodoxin unfolded in 6.0 M GuHCl behaves as a random coil. Far-UV CD shows that apoflavodoxin in 6.0 M GuHCl exhibits random coil behavior (Fig. 2.3a). This behavior is confirmed by ^{1}H-^{15}N R_2 relaxation rates (Fig. 2.4a), which are sensitive reporters of variations in polypeptide backbone dynamics.

The intrinsic properties (i.e., differences in side chain size) of an amino acid sequence play a significant role in determining sequence-dependent variations in backbone dynamics of an unfolded protein [23, 24]. Hence, the R_2 relaxation rates of apoflavodoxin in 6.0 M GuHCl are analyzed using a simple model that takes these variations in intrinsic flexibility of amino acid residues into account (Equation 2.1, Materials and Methods). The model describes the R_2 relaxation rates quite well (Fig. 2.4a). Thus, apoflavodoxin in 6.0 M GuHCl indeed behaves as a random coil and is devoid of regions with reduced flexibility.

Secondary shifts reveal three helical regions in unfolded apoflavodoxin in 3.4 M GuHCl.
Secondary shifts, which are deviations of observed chemical shifts from their random coil values, are good indicators for formation of secondary structure in proteins [25, 26]. To properly detect possible residual structure in unfolded proteins by using secondary shifts, corrections for sequence effects
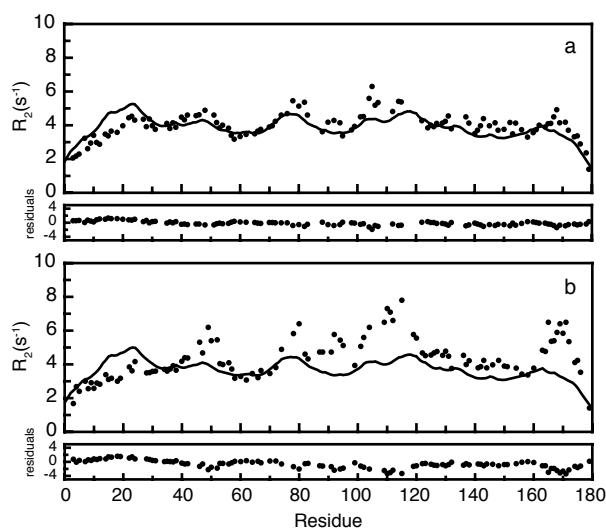
Figure 2.4 Sequence-dependent R_2 relaxation rates of unfolded apoflavodoxin. A model (Equation 2.1), which takes into account sequence-dependent variations in backbone motions due to differences in radius of gyration of amino acid residues, is fitted to the R_2 values obtained at 6.0 M (a) or 3.4 M GuHCl (b). A 700 MHz spectrometer was used.

should be taken into account because in particular $^{13}C'$, H^N and ^{15}N chemical shifts are highly sensitive to local amino acid sequence [27, 28]. To drastically improve the reliability for detecting residual structure within unfolded apoflavodoxin in 3.4 M GuHCl, secondary shifts are calculated using the chemical shifts of the random coil protein in 6.0 M GuHCl as intrinsic reference. Using this method, secondary shifts of amino acid residues are no longer influenced by the type of neighboring residues and thus become very sensitive reporters of secondary structure formation in unfolded apoflavodoxin.

Stretches of positive $^{13}C^\alpha$ and 13C' secondary shifts reveal α-helices, whereas stretches of negative $^{13}C^\alpha$ and $^{13}C'$ shifts indicate β-strands. Three regions of unfolded apoflavodoxin in 3.4 M GuHCl show contiguous positive $^{13}C^\alpha$ and $^{13}C'$ secondary shifts (Fig. 2.5). These α–helical regions comprise Ala41 – Ile52 (Helix-U1), Glu104 – Lys118 (Helix-U2) and Thr160 – Ala169 (Helix-U3).

Further support for presence of three helical regions in unfolded apoflavodoxin in 3.4 M GuHCl is obtained from H^N, $^{13}C^\beta$ and ^{15}N secondary shifts. Upon helix formation these secondary shifts should be negative [25], as is indeed observed for the majority of residues involved (Fig. 2.5). Note that several N-terminal residues of the helices show positive H^N secondary shifts. This effect is attributed to the helix dipole [29], which is a general property of helices and develops through alignment of amide groups with hydrogen-bonded carbonyl groups [30]. The observation that the three helices in unfolded apoflavodoxin have a dipole implies that these helices are cooperatively formed.

The transient nature of secondary structure elements in unfolded proteins causes their secondary shifts to be small. The secondary shift amplitudes of the helices detected in unfolded apoflavodoxin in 3.4 M GuHCl (Fig. 2.5) are one-tenth of those of fully formed regular α-helices in folded proteins, which have $^{13}C^\alpha$ secondary shifts of about 2 - 3 ppm [25]. This observation implies that the helices in unfolded apoflavodoxin in 3.4 M GuHCl are populated for about 10% of the time. Chemical shifts are the only means to detect such small populations, which cannot be detected by

methods based on NOEs or coupling constants [23]. Secondary shifts of $^{13}C^\beta$, ^{15}N and H^N reveal no apparent secondary structure outside these helical regions.

Far-UV CD confirms presence of helices in unfolded apoflavodoxin in 3.4 M GuHCl. The far-UV CD spectrum of unfolded apoflavodoxin in 3.4 M GuHCl suggests the presence of residual structure (Fig. 2.3a). The difference CD spectrum, obtained by subtracting the far-UV CD spectrum of apoflavodoxin in 6.0 M GuHCl from the one obtained in 3.4 M GuHCl, has a minimum at about 222 nm (Fig. 2.3b). This observation is typical for the presence of α-helices.

The molar ellipticity at 222 nm of native apoflavodoxin is -7.6×10^3 deg × cm²/dmol. A fully helical 179-residue protein has an ellipticity at 222 nm of -37.5×10^3 deg × cm²/dmol at 25 °C [31]. As the difference CD spectrum (Fig. 2.3b) shows a molar ellipticity at 222 nm of -0.79×10^3 deg × cm²/dmol, unfolded apoflavodoxin in 3.4 M GuHCl has a helical content of 2.1%.

NMR data show that the helical regions of unfolded apoflavodoxin in 3.4 M GuHCl comprise 37 residues, which is 21% of the total number of residues. Secondary shift amplitudes reveal that these regions populate the helical state during about 10% of the time. Thus, unfolded apoflavodoxin in 3.4 M GuHCl has a helical content of 2.1%, matching the value inferred from CD data.

Relaxation data show that four regions with restricted flexibility exist in unfolded apoflavodoxin in 3.4 M GuHCl. To further characterize unfolded apoflavodoxin in 3.4 M GuHCl, its dynamic features are probed by ^1H-^{15}N R_1, R_2 and NOE relaxation experiments, which reveal backbone flexibilities and conformational exchange processes on (sub)nano- to milliseconds timescales.

The ^1H-^{15}N NOE values of unfolded apoflavodoxin are distributed around 0.23 (Fig. 2.6a). The R_1 relaxation rates of apoflavodoxin in 3.4 M GuHCl are uniformly distributed around an average of 1.5 s^{-1} (Fig. 2.6b). Similar average R_1 values are obtained for other unfolded proteins [23, 24]. Both N- and C-terminal residues have even lower R_1 relaxation rates and negative NOE values as these residues are very flexible.

The R_2 relaxation rates of apoflavodoxin in 3.4 M GuHCl vary between about 2 and 8 s^{-1} (Fig. 2.6c) and are much less uniformly distributed than R_1 relaxation rates. Four stretches of residues have elevated R_2 values: Ala41 - Gly53, Glu72 - Gly83, Gln99 - Ala122, and Thr160 - Gly176, implying restricted flexibility on the (sub)nanosecond timescale in these regions of the unfolded protein.

The increased R_2 relaxation rates are not due to exchange between different conformers on micro- to millisecond timescales because similar rates are observed using 500 and 700 MHz spectrometers (Fig. 2.6c). If exchange occurs, its contribution to R_2 is proportional to the square of magnetic field strength involved. Without exchange R_2 is only marginally affected by changes in field strength [32]. In summary, four regions of apoflavodoxin in 3.4 M GuHCl exhibit reduced backbone flexibility on the (sub)nanosecond timescale.

Helix formation and hydrophobic interactions cause reduced flexibility in unfolded apoflavodoxin. Because four regions of unfolded apoflavodoxin have reduced flexibility on the (sub)nanosecond timescale, the simple model that takes into account sequence-dependent variations in backbone motions of an unfolded protein cannot describe the experimental R_2 data. Clearly, the elevated R_2 values of residues 41 – 53, 72 – 83, 99 – 122, and 160 – 176 are not fitted well using Equation 2.1 (Fig. 2.4b).

In the regions with restricted flexibility, interactions within and/or between clusters of residues most likely influence backbone motions of unfolded apoflavodoxin. Such type of interactions are also observed for urea-denatured apomyoglobin [23] and lysozyme [12]. The restricted flexibility is mainly due to helix formation in three of the four regions mentioned, as shown by analysis of secondary shifts of unfolded apoflavodoxin. The region with reduced flexibility that comprises residues 72 – 83 does not form α–helical secondary structure, as no positive secondary shifts of the corresponding $^{13}C^{α}$ and $^{13}C'$ nuclei are observed (Fig. 2.5). However, this region has negative H^N secondary shifts. Thus, residues 72 – 83 probably form transiently ordered structure that is neither α–helix nor β–strand.

The average area buried upon folding (AABUF) is correlated with hydrophobicity [33] and corresponds to sequence-dependent dynamic variations due to hydrophobic interactions in unfolded proteins [23, 34]. The four regions with restricted flexibility in unfolded apoflavodoxin match with regions of large AABUF (Figs. 2.7a,b). Hydrophobic interactions will stabilize initially formed structures within unfolded apoflavodoxin.

Transient non-native docking of ordered segments in unfolded apoflavodoxin. To probe via chemical shift changes whether the four ordered regions in unfolded apoflavodoxin in 3.4 M GuHCl interact with one another, use is made of site-directed mutagenesis to slightly change the hydrophobicity of one of the ordered regions. We replaced glutamine at position 48 by cysteine, which has hydrophobic characteristics at pH 6. In addition, this replacement enables site-directed spin labeling of the protein, allowing future paramagnetic relaxation enhancement experiments.
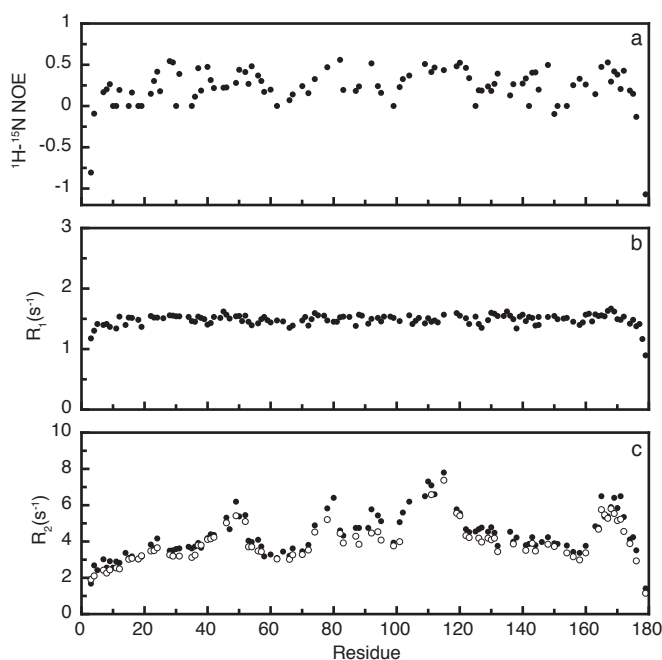
Figure 2.6 Sequence-dependent relaxation data of apoflavodoxin unfolded in 3.4 M GuHCl. Data obtained on a 700 MHz spectrometer. (•) (a) ^{1}H-^{15}N heteronuclear NOEs, (b) R_1 relaxation rates, and (c) R_2 relaxation rates (data obtained on a 500 MHz spectrometer (o)).
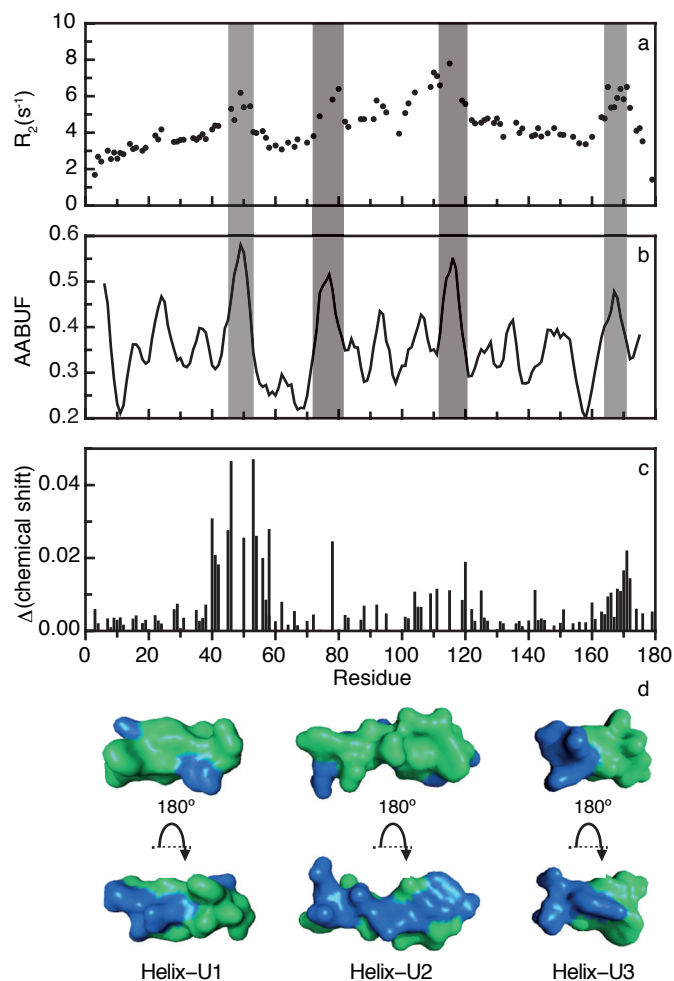
Figure 2.7 The four regions with elevated R_2 relaxation rates in unfolded apoflavodoxin in 3.4 M GuHCl correspond to regions with large surface area buried upon folding. (a) R_2 values of apoflavodoxin unfolded in 3.4 M GuHCl. (b) Average area buried upon folding (AABUF). Gray boxes highlight maxima in AABUF. (c) Differences between backbone amide chemical shifts of 48Q- and 48C-apoflavodoxin variants unfolded in 3.4 M GuHCl. Chemical shift differences are calculated using Δ(chemical shift) $= |\Delta\delta H^N| + 0.1 \cdot |\Delta\delta^{15}N|$. (d) Cartoon drawing of the helical regions in unfolded apoflavodoxin. Hydrophobic residues are shown in green.

Residue 48 resides in the middle of Helix-U1 in unfolded apoflavodoxin. Figure 2.7c shows that chemical shifts of many residues in unfolded apoflavodoxin change due to introducing this cysteine. Interestingly, several residues of Helix-U2 and Helix-U3 are also affected. Furthermore, Leu78, which resides in the region with reduced flexibility comprising residues 72 – 83, shows chemical shift changes. Interaction of Helix-U1 with Leu78 explains the reduced flexibility and the contiguous stretch of negative H^N secondary shifts for residues 75 - 81 (Fig. 2.5a). The signs of the H^N and ^{15}N chemical shift changes upon replacing Gln48 by Cys48 reveal that these changes are due to increased population of all three helices observed in unfolded apoflavodoxin (Fig. S2.1, Supporting Information). Stabilization of Helix-U2 and Helix-U3 can only occur if long-range hydrophobic interactions between these helices and Helix-U1 exist in unfolded apoflavodoxin. The chemical shift changes show that Helix-U1 has transient hydrophobic interactions with all other ordered regions in the unfolded protein and implies non-native docking of these regions.

Discussion

The most interesting and physiologically relevant situation to characterize unfolded apoflavodoxin would be under native conditions. Unfortunately, the overwhelming majority of protein molecules are then natively folded (Fig. 2.1b). Detailed characterization of unfolded apoflavodoxin by NMR spectroscopy thus occurred in presence of a chemical denaturant. The corresponding unfolded state may therefore differ from the unfolded state under native conditions.

Four regions of unfolded apoflavodoxin in 3.4 M GuHCl have reduced flexibility, which is due to transient helix formation and local and non-local hydrophobic interactions (Figs. 2.8a,b). Residues 41 to 45 of Helix-U1 and residues 108 to 118 of Helix-U2 and all residues of Helix-U3 are helical in the native protein as well. These stretches of residues thus have intrinsic propensity to form native-like helices. Helices are the only regular secondary structure elements in unfolded apoflavodoxin in 3.4 M GuHCl and these helices are sufficiently stable to be present during about 10% of the time.

Non-native structure elements in unfolded apoflavodoxin. Remarkably, the C-terminal part (residues 47 to 52) of Helix-U1 in unfolded apoflavodoxin is non-native. In native protein, several of these residues form the N-terminal part of β-strand-3, which is the central strand of apoflavodoxin's parallel β–sheet (Fig. 2.8a). The N-terminal part (residues 104 to 107) of Helix-U2 in unfolded apoflavodoxin also is non-native, as it is not helical in native protein. Residues 72 – 83 (light blue in Fig. 2.8a) form transiently ordered structure in unfolded apoflavodoxin that is neither α–helix nor β–strand and is involved in non-native hydrophobic interactions with Helix-U1.

Residues Ala41 – Ile52 (i.e., Helix-U1) form a transient helix in unfolded apoflavodoxin. However, part of this helix (residues 49 to 52) becomes β-strand-3 in native protein. To further investigate the intrinsic propensity for secondary structure type formation of this region, use is made of Scratch Protein Predictor (SSPro) [35], Jnet [36] and AGADIR [37]. These programs use structure prediction algorithms that are based solely on propensity of individual residues to be part of regular secondary structure types. Although no direct proof, these algorithms predict helix formation in a peptide fragment comprising residues Val36 to Glu61 of apoflavodoxin. The helix begins either at Arg38 (AGADIR), Val39 (Jnet) or Ala41 (SSPro) and continues until Ile51 (AGADIR, Jnet, SSPro). AGADIR assigns positive $^{13}C^\alpha$ secondary shifts to residues Glu42 to Leu52 (results not shown), matching our experimental data (Fig. 2.5c). Formation of Helix-U1 in unfolded apoflavodoxin is an

intrinsic property of its amino acid sequence. Helix-U1 needs not to be induced through contacts with other regions of the unfolded protein. These observations strongly suggest that formation of native β–strand-3 during apoflavodoxin folding requires specific tertiary contacts.

Structure formation in unfolded apoflavodoxin promotes off-pathway intermediate production. The chemical shift changes upon replacing Gln48 by Cys48 show that Helix-U1 has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin and imply transient non-native docking of these regions.

Support for conformational pre-organization in unfolded apoflavodoxin promoting off-pathway intermediate formation is provided by hydrogen/deuterium (H/D) exchange results. Detection by NMR spectroscopy of native state H/D exchange in presence of small amounts of a denaturant identified five unfolding clusters of residues within native apoflavodoxin [20]. Four of these clusters unfold subglobally in a cooperative manner. The resulting conformations are partially unfolded forms (PUFs) of the protein. Both PUF1 and PUF2 are unfolding excursions that start from native apoflavodoxin but do not continue to the unfolded state and do not reside on the productive folding route. In contrast, both PUF3 and PUF4 (Fig. 2.8c) almost certainly are PUFs of the off-pathway folding intermediate [20] and are positioned between this species and apoflavodoxin's unfolded state. Remarkably, the study presented here reveals that the residues that are folded and protected against H/D exchange in PUF3 are characterized by enhanced R_2 relaxation rates and by formation of transient secondary structure within apoflavodoxin in 3.4 M GuHCl (Fig. 2.8b). These ordered structures are populated about 10% of the time and are distinctively different from the random coil structures that are populated during the largest fraction of the time. The four regions with transient structure in unfolded apoflavodoxin actually reflect population of PUF3, whereas Helix-U2 reflects population of PUF4. As both PUFs are excursions from the off-pathway species, secondary structure formation in unfolded apoflavodoxin indeed promotes non-productive folding toward an off-pathway folding intermediate.

Implications for protein folding. The presence of non-native secondary structure elements in unfolded proteins is probably a widespread phenomenon. However, subsequent formation of folding intermediates that contain these non-native structure elements is likely but rarely reported. For example, in case of β–lactoglobulin, an on-pathway intermediate with some fluctuating non-native helical structure near the N-terminus, which is later converted to β–sheet, is formed [38]. In case of the N-domain of phosphoglycerate kinase, a non-native helical species is formed, but it is unknown whether this species is on- or off-pathway [39].

An off-pathway intermediate plays a major role during A. vinelandii apoflavodoxin folding and is also observed during the kinetic folding of three other proteins with an α-β parallel topology of which the folding mechanism has been studied [22]: *Anabaena* apoflavodoxin [40], *Fusarium solani pisi* cutinase [41] and *Escherichia coli* CheY [42]. In this study, it is proven for the first time that formation of native and non-native helices within an unfolded α-β parallel protein and subsequent helix docking leads to formation of a compact helical off-pathway intermediate. Three relatively stable helices are populated in unfolded apoflavodoxin and dock non-natively onto one another (Fig. 2.7c) through transient interactions between their hydrophobic sides (Fig. 2.7d). The non-native docking of these helices prevents formation of the parallel β-sheet of native protein. To produce native α-β parallel protein molecules, the off-pathway species needs to unfold and as a result non-native interactions and non-native secondary structure are disrupted.
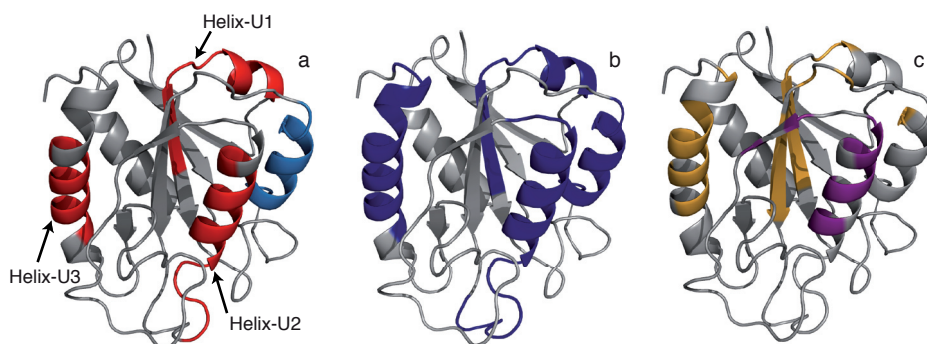
Figure 2.8 Cartoon drawings of flavodoxin from *A. vinelandii* (pdb ID 1YOB). The FMN cofactor is not shown. (a) Regions of unfolded apoflavodoxin in 3.4 M GuHCl that adopt α-helical structure are colored red. Part of β-strand-3 of native flavodoxin has helical characteristics in the unfolded protein (i.e., Helix-U1). The region that forms transiently ordered structure that is neither α–helix nor β–strand and has non-native hydrophobic interactions with Helix-U1 is colored light blue. (b) Regions of unfolded apoflavodoxin in 3.4 M GuHCl with elevated R_2 relaxation rates are colored blue. (c) Partially unfolded forms PUF3 and PUF4 of apoflavodoxin detected by native-state H/D exchange [20]. The clusters of cooperatively unfolding residues are shown in orange and purple. In PUF3 (orange and purple) the amides of both clusters are protected against exchange and the remainder of the protein is unfolded. In PUF4 only the amides of the purple cluster are protected and all other parts of the protein are unfolded.

The presented study shows that acquisition of native-like topology is not necessarily the general result of the initial collapse in protein folding. Rather than directing productive folding, conformational pre-organization in the unfolded state of an α-β parallel type protein promotes off-pathway species formation. During kinetic folding of this type of protein, helices are formed much more easily and thus more rapidly than sheets, especially as a parallel β-sheet is involved. Formation of helices can occur on the nanosecond timescale [43]. This is due to the highly local character of interactions in helices, whereas the residues that have to be brought into contact to form a parallel β-sheet are separated by many residues from one another [43]. Our observations indicate that especially proteins that contain domains with an α-β parallel topology seem susceptible to off-pathway intermediate formation.

Supporting Information

Additional support for the presence of hydrophobic tertiary interactions between residual helices in unfolded apoflavodoxin. Chemical shift changes due to amino acid substitutions in unfolded polypeptides only extend to about four neighboring residues on each side of the substituted residue, as is observed for OmpX [52], lysozyme [53], and Im7 [54]. All these substitutions were made in regions of unfolded proteins that form clusters of hydrophobic structure, just as is the case for the mutation at position 48 in unfolded apoflavodoxin. Consequently, it is remarkable that in unfolded apoflavodoxin, replacement of Gln48 by Cys48 affects the chemical shifts of much more than eight neighboring residues (i.e., residues 40 to 58) (Fig. 2.7c). In addition, the chemical shifts of many other residues not sequentially near residue 48 are also affected, although the magnitude of these chemical shift changes is smaller than observed for the direct sequential neighbors of Cys48 (Fig. 2.7c). We ascribe the chemical shift changes observed to Helix-U1 having transient hydrophobic interactions with all other ordered regions in the unfolded protein. Additional support for this conclusion is given in the following.
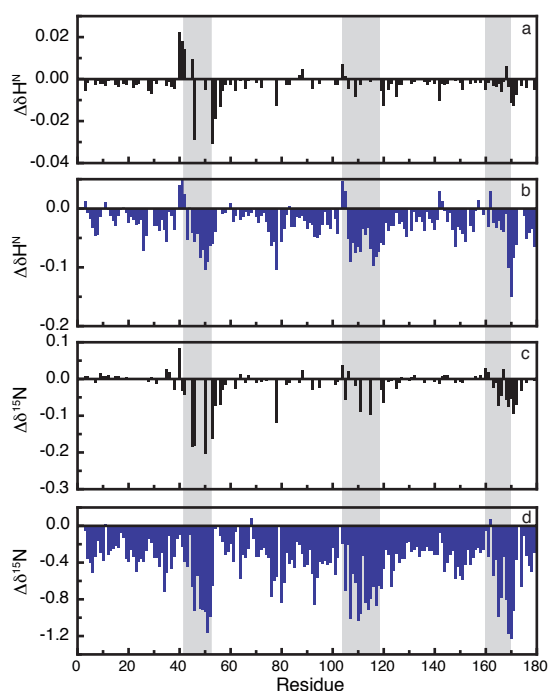
Figure S2.1 Upon replacing Q48 by C48 the helices in unfolded apoflavodoxin in 3.4 M GuHCl become more heavily populated. Differences between backbone amide chemical shifts of 48C- and 48Q-apoflavodoxin variants unfolded in 3.4 M GuHCl are shown; (a) $\Delta\delta H^N$ and (c) $\Delta\delta^{15}N$. Chemical shift differences are calculated using: $\Delta\delta H^N = \delta H^N C48 - \delta H^N Q48$; $\Delta\delta^{15}N = \delta^{15}NC48 - \delta^{15}NQ48$. For comparison, secondary shifts of unfolded 48Q-apoflavodoxin in 3.4 M GuHCl, as calculated by intrinsic referencing, are shown for (b) H^N and for (d) ^{15}N. The regions of apoflavodoxin that form α-helical structure in 3.4 M GuHCl (i.e., Helix-U1, Helix-U2 and Helix-U3) are highlighted.

Figures S2.1a and S2.1c show chemical shift changes of the backbone amides of unfolded apoflavodoxin in 3.4 M GuHCl upon replacing Q48 by C48. For comparison, secondary shifts of the backbone amides of unfolded 48Q-apoflavodoxin in 3.4 M GuHCl, as shown in Figures 2.5a and 2.5e, are also included in Figures S2.1b and S2.1d.

Comparison of Figure S2.1a with Figure S2.1b shows that replacement of Gln48 by Cys48 leads to a pattern of H^N chemical shift differences that is roughly comparable to the secondary shift pattern obtained of unfolded 48Q-apoflavodoxin in 3.4 M GuHCl (note for example the characteristic positive chemical shifts differences for the N-terminal part of Helix-U1). A similar observation is made for the ^{15}N chemical shift differences (compare Figure S2.1c with Figure S2.1d). As secondary shifts reveal the presence of helices in unfolded apoflavodoxin in 3.4 M GuHCl, the chemical shift changes observed in Figures S2.1a and S2.1c imply increased population of Helix-U1 upon replacing Gln48. Although not as high in amplitude as observed for Helix-U1, Figures S2.1a and S2.1c show that Helix-U2 and Helix-U3 also become more populated. The data shown in Figure S2.1 demonstrate that increasing the hydrophobicity at position 48 through replacement of glutamine by cysteine leads to increased population of helices in three separate regions of the unfolded protein. This observation is additional proof that long-range hydrophobic interactions must exist in the unfolded protein.

Backbone resonance assignments of unfolded apoflavodoxin

Table S2.1 Backbone resonance assignment of *A. vinelandii* apoflavodoxin in 6.0 M GuHCl

Residue		HN	¹³C'	¹³Cα	¹³Cβ	¹⁵N
1	ala		174.269	52.019	19.928	
2	lys	8.595	176.796	56.551		122.309
3	ile	8.176	176.989	61.399	39.143	122.499
4	gly	8.393	174.158	45.487		113.499
5	leu	7.985	177.279	55.328	42.801	122.526
6	phe	8.168	175.997	57.617	40.04	120.913
7	phe	8.246	176.635	57.955	40.042	122.468
8	gly	8.062	174.537	45.627		110.508
9	ser	8.241	175.08	58.529	64.246	116.209
10	asn	8.625	176	53.492	39.11	121.604
11	thr	8.067	175.674	62.003	70.119	113.342
12	gly	8.399	174.548	45.6		111.692
13	lys	8.172	177.429	56.657	33.151	121.446
14	thr	8.184	175.096	61.917	70.227	115.847
15	arg	8.381	176.651	56.219	31.217	124.389
16	lys	8.385	177.034	56.569	33.335	124.076
17	val	8.123	176.332	62.053	33.182	121.887
18	ala	8.346	178.094	52.533	19.757	128.511
19	lys	8.296	177.153	56.645	33.42	121.808
20	ser	8.321	175.019	58.21	64.05	117.647
21	ile	8.157	176.637	61.181	38.917	123.197
22	lys	8.298	176.913	56.455	33.28	126.002
23	lys	8.303	176.803	56.511	33.405	124.046
24	arg	8.29	176.542	56.271	31.37	122.968
25	phe	8.322	176.123	57.699	40.055	121.54
26	asp	8.436	176.328	54.421	41.638	122.221
27	asp	8.266	176.977	54.46	41.637	120.87
28	glu	8.443	177.395	57	30.399	121.297
29	thr	8.215	175.451	62.207	70.123	114.563
30	met	8.34	176.862	55.799	33.225	122.832
31	ser	8.262	174.971	58.612	64.093	117.087
32	asp	8.429	176.567	54.437	41.602	123.252
33	ala	8.132	178.128	52.784	19.711	123.901
34	leu	8.144	177.7	55.474	42.629	121.144
35	asn	8.371	175.823	53.313	39.084	120.159
36	val	7.962	176.277	62.424	33.013	119.508
37	asn	8.44	175.594	53.398	39.311	122.082
38	arg	8.204	176.759	56.34	30.996	122.038
39	val	8.143	176.713	62.448	33.04	121.321
40	ser	8.383	174.983	58.252	64.281	119.898
41	ala	8.35	178.371	53.125	19.717	126.509
42	glu	8.318	176.817	56.839	30.355	119.321
43	asp	8.185	176.583	54.489	41.568	121.017
44	phe	8.023	176.203	57.861	39.882	120.414
45	ala	8.194	178.035	52.692	19.697	124.99
46	gln	8.168	176.331	56.302	29.709	119.576
47	tyr	8.039	176.356	57.774	38.94	120.792
48	gln	8.242	175.998	56.275	29.721	121.796
49	phe	8.032	175.958	57.705	39.733	120.847
50	leu	8.146	177.277	55.318	42.824	123.883
51	ile	8.129	176.635	61.193	38.594	122.968
52	leu	8.29	178.208	55.248	42.627	126.825
53	gly	8.324	174.293	45.358		110.183
54	thr	8.07	173.62	59.927	69.854	115.869
55	pro		177.547	63.508	32.28	
56	thr	8.221	175.237	61.906	70.144	115.024
57	leu	8.268	178.258	55.517	42.652	124.897
58	gly	8.489	174.712	45.462		110.608
59	glu	8.331	177.615	56.96	30.58	121.13
60	gly	8.465	174.407	45.445		110.295
61	glu	8.175	176.853	56.325	30.837	120.892
62	leu	8.304	175.842	53.277	41.774	124.229
63	pro		177.962	63.473	32.313	
64	gly	8.432	174.523	45.512		109.93
65	leu	8.119	178.186	55.31	42.859	122.247
66	ser	8.368	175.344	58.536	64.115	117.272
67	ser	8.38	175.03	58.65	64.095	118.078
68	asp	8.394	176.641	54.635	41.53	122.801
69	ala	8.1	178.408	52.945	19.73	123.77

#	res					
70	glu	8.307	176.936	56.839	30.322	119.717
71	asn	8.267	175.785	53.424	39.324	119.322
72	glu	8.362	177.017	56.749	30.589	121.467
73	ser	8.313	175.011	58.678	64.083	116.783
74	trp	8.106	176.742	57.617	29.954	123.431
75	glu	8.156	176.61	56.642	30.553	122.205
76	glu	8.199	176.42	56.66	30.68	121.881
77	phe	8.167	175.66	57.481	39.974	121.384
78	leu	8.13	175.192	52.76	42.109	125.74
79	pro		177.245	63.018	32.285	
80	lys	8.45	177.164	56.776	33.061	123.117
81	ile	8.167	176.584	61.193	38.917	122.905
82	glu	8.458	177.288	56.702	30.682	125.474
83	gly	8.364	174.422	45.61		110.607
84	leu	8.055	177.464	55.326	42.853	121.946
85	asp	8.347	176.463	54.185	41.634	121.113
86	phe	8.196	176.467	57.905	39.767	121.364
87	ser	8.36	175.465	58.774	64.256	117.756
88	gly	8.09	174.412	45.618		111.145
89	lys	8.172	177.4	56.657	33.515	121.446
90	thr	8.207	175.119	61.959	70.206	116.109
91	val	8.096	176.094	62.096	33.137	122.534
92	ala	8.258	177.719	52.391	19.723	127.883
93	leu	8.095	177.486	55.318	42.8	122.472
94	phe	8.138	176.793	57.656	40.057	120.251
95	gly	8.393	174.593	45.638		110.825
96	leu	8.148	178.496	55.578	42.641	122.065
97	gly	8.431	174.523	45.63		109.93
98	asp	8.261	176.821	54.46	41.637	121.083
99	gln	8.337	176.417	56.11	29.649	120.439
100	val	8.008	176.936	62.481	33.041	120.473
101	gly	8.309	173.774	45.215		112.52
102	tyr	8.082	174.636	56.157	38.484	121.224
103	pro		177.39	63.372	32.189	
104	glu	8.554	176.871	57.288	30.241	121.351
105	asn	8.345	175.608	53.305	39.197	118.88
106	tyr	8.06	176.386	58.478	38.955	120.967
107	leu	8.091	177.663	55.525	42.451	123.037
108	asp	8.185	176.583	54.489	41.568	121.059
109	ala	8.093	178.28	52.94	19.74	123.889
110	leu	8.126	178.502	55.668	42.539	121.052
111	gly	8.264	174.637	45.651		109.45
112	glu	8.216	177.013	56.585	30.598	121.026
113	leu	8.19	177.538	55.55	42.544	122.963
114	tyr	8.087	176.423	57.978	39.149	120.749
115	ser	8.12	174.55	58.233	64.052	117.26
116	phe	8.14	176.075	58.271	39.957	122.802
117	phe	8.141	176.233	57.949	39.951	121.092
118	lys	8.176	176.545	56.664	33.341	122.913
119	asp	8.278	176.774	54.378	41.624	121.837
120	arg	8.283	177.3	56.645	30.947	121.858
121	gly	8.432	174.259	45.512		109.868
122	ala	8.083	178.191	52.54	19.82	124.084
123	lys	8.287	176.935	56.645	33.137	121.664
124	ile	8.172	177.75	61.148	38.75	123.422
125	val	8.15	176.915	62.389	33.085	124.856
126	gly	8.314	174.442	45.365		113.116
127	ser	8.124	174.914	58.379	64.042	116.139
128	trp	8.177	176.809	57.64	30.042	123.291
129	ser	8.136	175.301	58.332	64.2	117.082
130	thr	8.129	174.987	61.892	69.936	115.332
131	asp	8.322	177.042	54.598	41.696	122.338
132	gly	8.211	174.35	45.596		109.195
133	tyr	8.062	176.54	58.156	39.218	120.608
134	glu	8.406	176.549	56.829	30.401	122.828
135	phe	8.04	176.356	57.898	39.733	120.739
136	glu	8.364	176.826	56.548	30.802	122.554
137	ser	8.292	175.305	58.389	64.175	117.1
138	ser	8.409	175.12	58.766	64.083	118.518
139	glu	8.39	176.641	56.607	30.553	122.736
140	ala	8.19	178.035	52.692	19.697	124.819
141	val	8	176.715	62.404	32.929	119.985
142	val	8.119	176.233	62.262	33.072	123.62
143	asp	8.389	177.019	54.481	41.739	124.569
144	gly	8.25	174.435	45.651		109.435
145	lys	8.105	176.801	56.826	33.429	121.362

146	phe	8.305	176.287	57.699	39.667	121.495	163	arg	8.365	177.022	56.548	30.802	122.467
147	val	7.984	176.758	62.393	33.093	122.076	164	val	8.037	176.392	62.3	33.02	121.33
148	gly	8.065	174.269	45.548		112.426	165	ala	8.223	178.211	52.629	19.29	127.636
149	leu	8.083	177.678	55.318	42.8	122.428	166	ala	8.214	178.357	53.43	19.398	123.776
150	ala	8.352	178.069	52.607	19.379	125.479	167	trp	7.867	176.834	57.718	29.433	119.007
151	leu	8.134	177.608	55.497	42.683	121.826	168	leu	7.756	177.459	55.14	42.566	123.186
152	asp	8.389	176.735	54.242	41.224	121.221	169	ala	7.957	178.23	52.793	19.643	124.17
153	leu	8.047	177.784	55.481	42.523	122.315	170	gln	8.194	176.386	56.117	29.736	119.598
154	asp	8.354	176.663	54.515	41.518	120.712	171	ile	7.962	176.051	61.007	39.135	121.364
155	asn	8.29	175.991	53.465	39.23	119.334	172	ala	8.277	175.891	50.815	18.592	128.942
156	gln	8.388	176.698	56.275	29.603	120.613	173	pro		177.354	63.302	32.282	
157	ser	8.308	175.586	58.678	64.083	116.788	174	glu	8.482	176.716	57.012	30.331	121.201
158	gly	8.429	174.571	45.614		111.444	175	phe	8.105	176.821	57.81	40.047	120.256
159	lys	8.188	177.498	56.657	33.515	121.393	176	gly	8.34	174.372	45.631		110.592
160	thr	8.183	174.823	61.836	70.137	115.195	177	leu	8.063	177.927	55.326	42.853	122.038
161	asp	8.421	176.652	54.437	41.602	123.314	178	ser	8.338	174.483	58.31	64.029	117.847
162	glu	8.356	177.017	56.749	30.589	121.714	179	leu	7.999	175.286	57.284	43.146	130.186

Table S2.2 Backbone resonance assignment of *A. vinelandii* apoflavodoxin in 3.4 M GuHCl

Residue		H^N	$^{13}C'$	$^{13}C^\alpha$	$^{13}C^\beta$	^{15}N							
1	ala						18	ala	8.348	177.578	52.541	19.623	128.384
2	lys		176.29	56.469	33.253		19	lys	8.288	176.619	56.674	33.38	121.45
3	ile	8.188	176.471	61.354	39.003	122.446	20	ser	8.296	174.465	58.193	63.939	117.253
4	gly	8.386	173.691	45.416		113.145	21	ile	8.142	176.142	61.221	38.877	122.933
5	leu	7.967	176.787	55.329	42.614	122.122	22	lys	8.278	176.379	56.443	33.242	125.669
6	phe	8.135	175.403	57.605	39.863	120.403	23	lys	8.276	176.271	56.465	33.264	123.546
7	phe	8.199	176.036	57.909	39.87	122.138	24	arg	8.264	176.035	56.326	31.302	122.579
8	gly	8.017	174.016	45.518		110.35	25	phe	8.302	175.517	57.761	39.722	121.178
9	ser	8.226	174.536	58.469	64.128	115.914	26	asp	8.364	175.761	54.337	41.45	121.827
10	asn	8.624	175.481	53.52	38.95	121.222	27	asp	8.219	176.504	54.456	41.476	120.601
11	thr	8.078	175.162	62.068	70.021	113.362	28	glu	8.442	176.906	57.156	30.297	121.153
12	gly	8.396	174.083	45.511		111.383	29	thr	8.214	174.958	62.422	69.919	114.381
13	lys	8.159	176.899	56.589	33.123	121.176	30	met	8.311	176.343	55.847	32.939	122.491
14	thr	8.163	174.478	61.992	70.058	115.6	31	ser	8.233	174.482	58.712	63.975	116.737
15	arg	8.352	176.095	56.208	31.189	124.176	32	asp	8.391	176.078	54.52	41.436	122.811
16	lys	8.373	176.475	56.415	33.249	123.84	33	ala	8.113	177.691	52.9	19.524	123.614
17	val	8.12	175.764	62.088	33.11	121.811	34	leu	8.107	177.185	55.457	42.552	120.428
							35	asn	8.314	175.304	53.298	38.946	119.645

#	aa					
36	val	7.956	175.779	62.504	32.819	119.399
37	asn	8.427	175.003	53.434	39.231	121.615
38	arg	8.167	176.187	56.349	31.058	121.699
39	val	8.143	176.145	62.444	32.966	121.185
40	ser	8.423	174.507	58.175	64.213	119.803
41	ala	8.398	178.122	53.264	19.427	126.279
42	glu	8.342	176.439	57.125	30.11	118.927
43	asp	8.131	176.196	54.788	41.34	120.573
44	phe	8.026	175.804	58.045	39.706	120.316
45	ala	8.153	177.786	53.096	19.294	124.324
46	gln	8.111	176.003	56.488	29.402	118.656
47	tyr	7.996	175.891	58.294	38.789	120.253
48	gln	8.158	175.605	56.504	29.43	120.896
49	phe	7.961	175.591	57.926	39.503	119.933
50	leu	8.041	176.898	55.46	42.704	122.948
51	ile	8.038	176.182	61.279	38.533	121.804
52	leu	8.226	177.694	55.289	42.596	125.841
53	gly	8.264	173.722	45.298		109.533
54	thr	8.03	173.014	59.907	69.817	115.837
55	pro		176.975	63.457		
56	thr	8.222	174.63	61.941	70.073	114.912
57	leu	8.259	177.724	55.51	42.578	124.721
58	gly	8.482	174.181	45.387		110.292
59	glu	8.33	177.077	56.875	30.452	120.848
60	gly	8.474	173.849	45.387		110.101
61	glu	8.153	176.253	56.257	30.809	120.499
62	leu	8.292	175.283	53.205	41.808	124.17
63	pro		177.437	63.533		
64	gly	8.413	174.035	45.496		109.355
65	leu	8.105	177.658	55.28	42.749	121.929
66	ser	8.362	174.788	58.438	64.024	116.929
67	ser	8.368	174.494	58.499	63.979	117.831
68	asp	8.386	176.083	54.52	41.436	122.886
69	ala	8.097	177.944	52.9	19.524	123.636
70	glu	8.293	176.415	56.858	30.195	119.386
71	asn	8.242	175.266	53.483	39.208	118.991
72	glu	8.35	176.493	56.753	30.365	121.213
73	ser	8.305	174.519	58.738	63.97	116.551
74	trp	8.083	176.278	57.662	29.736	123.121
75	glu	8.116	176.152	56.79	30.414	121.859
76	glu	8.136	175.887	56.842	30.4	121.112
77	phe	8.11	175.081	57.51	39.87	120.784
78	leu	8.025	174.621	52.829	42.12	125.189
79	pro		176.683	62.997		
80	lys	8.394	176.64	56.689	32.877	122.285
81	ile	8.133	176.041	61.191	38.931	122.283
82	glu	8.448	176.739	56.819	30.519	125.181
83	gly	8.369	173.961	45.491		110.198
84	leu	8.023	176.858	55.269	42.753	121.581
85	asp	8.315	175.923	54.155	41.423	120.66
86	phe	8.182	175.915	58.017	39.382	121.189
87	ser	8.344	174.949	58.818	64.052	117.556
88	gly	8.078	173.985	45.564		110.844
89	lys	8.151	176.85	56.689	33.134	121.002
90	thr	8.172	174.56	61.992	70.058	115.684
91	val	8.072	175.603	62.285	32.957	122.28
92	ala	8.231	177.272	52.453	19.55	127.366
93	leu	8.046	176.991	55.346	42.565	121.61
94	phe	8.088	176.203	57.703	39.884	119.69
95	gly	8.348	174.089	45.565		110.432
96	leu	8.12	177.972	55.531	42.581	121.651
97	gly	8.425	173.998	45.496		109.551
98	asp	8.233	176.249	54.456	41.476	120.657
99	gln	8.302	175.864	56.025	29.484	120.036
100	val	8.005	176.364	62.511	32.85	120.231
101	gly	8.291	173.227	45.112		111.987
102	tyr	8.039	174.041	56.105	38.525	120.88
103	pro		176.913	63.251		
104	glu	8.601	176.49	57.578	30.044	121.154
105	asn	8.374	175.289	53.542	38.917	118.179
106	tyr	8.009	175.875	58.812	38.847	120.765
107	leu	8	177.495	55.791	42.316	122.024
108	asp	8.127	176.448	54.788	41.34	120.54
109	ala	8.017	178.08	53.218	19.399	123.272
110	leu	8.061	178.084	55.913	42.274	120.025
111	gly	8.19	174.472	45.809		108.49

112	glu	8.182	176.788	56.84	30.362	120.704
113	leu	8.145	177.222	55.908	42.483	122.124
114	tyr	8.054	176.056	58.359	38.966	119.841
115	ser	8.051	174.264	58.636	63.755	116.47
116	phe	8.043	175.674	58.475	39.71	122.138
117	phe	8.059	175.796	58.174	39.71	120.225
118	lys	8.102	176.104	56.81	33.146	122.261
119	asp	8.219	176.252	54.429	41.469	121.153
120	arg	8.221	176.793	56.662	30.777	121.394
121	gly	8.408	173.772	45.496		109.612
122	ala	8.045	177.611	52.56	19.818	123.772
123	lys	8.258	176.404	56.545	33.151	121.001
124	ile	8.142	176.171	61.221	38.877	122.933
125	val	8.143	176.355	62.457	32.995	124.551
126	gly	8.288	173.891	45.314		112.715
127	ser	8.101	174.304	58.321	63.969	115.781
128	trp	8.142	176.171	57.593	29.93	122.933
129	ser	8.087	174.776	58.315	64.084	116.897
130	thr	8.126	174.465	62.008	69.788	115.163
131	asp	8.284	176.472	54.665	41.598	122.054
132	gly	8.205	173.879	45.519		108.935
133	tyr	8.04	175.933	58.242	39.239	120.381
134	glu	8.391	175.942	56.689	30.345	122.494
135	phe	8.018	175.766	58.045	39.706	120.472
136	glu	8.326	176.253	56.639	30.574	122.288
137	ser	8.286	174.765	58.362	64.066	116.875
138	ser	8.392	174.569	58.78	64.009	118.245
139	glu	8.366	176.076	56.691	30.424	122.463
140	ala	8.174	177.454	52.648	19.57	124.504
141	val	7.981	176.196	62.388	32.867	119.67
142	val	8.148	175.717	62.304	32.997	123.55
143	asp	8.401	176.486	54.455	41.545	124.381
144	gly	8.246	173.996	45.719		108.94
145	lys	8.082	176.247	56.689	33.222	120.956
146	phe	8.297	175.766	57.761	39.722	121.177
147	val	7.947	176.17	62.478	33.08	121.808
148	gly	8	173.767	45.482		111.846
149	leu	8.052	177.093	55.48	42.635	121.989
150	ala	8.316	177.532	52.631	19.249	124.935
151	leu	8.091	177.056	55.471	42.61	121.237
152	asp	8.332	176.236	54.237	41.09	120.798
153	leu	8.048	177.33	55.48	42.635	122.053
154	asp	8.325	176.121	54.658	41.401	120.292
155	asn	8.251	175.481	53.483	39.208	118.991
156	gln	8.384	176.202	56.283	29.531	120.351
157	ser	8.322	175.071	58.738	63.97	116.551
158	gly	8.428	174.118	45.519		111.154
159	lys	8.177	176.992	56.59	33.306	121.137
160	thr	8.18	174.263	61.972	69.996	115.144
161	asp	8.389	176.25	54.52	41.436	122.977
162	glu	8.386	176.681	57.098	30.345	121.785
163	arg	8.341	176.746	56.782	30.756	121.93
164	val	8.004	176.053	62.811	32.783	120.964
165	ala	8.202	178.08	53.014	19.063	126.652
166	ala	8.178	178.207	53.576	19.105	122.995
167	trp	7.827	176.645	58.07	29.185	118.667
168	leu	7.747	177.184	55.538	42.502	122.386
169	ala	7.856	177.815	53.002	19.357	122.999
170	gln	8.043	175.944	56.1	29.556	118.374
171	ile	7.878	175.511	61.049	38.938	120.445
172	ala	8.216	175.358	50.833	18.439	128.57
173	pro		176.971	63.347	32.237	
174	glu	8.481	176.253	57.04	30.123	120.537
175	phe	8.093	176.283	57.92	39.785	120.15
176	gly	8.288	173.875	45.575		110.265
177	leu	8.021	177.298	55.269	42.753	121.604
178	ser	8.304	173.712	58.193	63.939	117.351
179	leu	7.933	174.633	57.098	43.219	129.891

References

1. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25**, 331-339.
2. Fersht, A. R. & Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell* **108**, 573-582.
3. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641-645.
4. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis. *Proteins* **21**, 167-195.
5. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S. & Plaxco, K. W. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* **101**, 12491-12496.
6. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem Rev* **104**, 3607-3622.
7. Mittag, T. & Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3-14.
8. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* **101**, 12497-12502.
9. Mok, K. H., Kuhn, L. T., Goez, M., Day, I. J., Lin, J. C., Andersen, N. H. & Hore, P. J. (2007). A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* **447**, 106-109.
10. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.
11. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487-489.
12. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.
13. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053-1056.
14. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208.
15. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366.
16. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.
17. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of Azotobacter vinelandii holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Prot Sci* **7**, 306-317.
18. Bollen, Y. J., Sanchez, I. E. & van Mierlo, C. P. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475-10489.
19. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem.* **280**, 7836-7844.
20. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci USA* **103**, 4095-4100.
21. van Mierlo, C. P. M, van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Prot Sci* **9**, 145-157.
22. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.
23. van Mierlo, C. P. M., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of Azotobacter vinelandii apoflavodoxin II occurs via a relatively stable folding intermediate. *Prot Sci* **7**, 2331-2344.

24. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. In *Meth enzymol* (Hirs, C. H. W. & Timasheff, S. N., eds.), Vol. 26, pp. 43-50. Academic Press, New York.

25. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M. & Skelton, N. J. (2007). *Protein NMR Spectroscopy: principles and practice*, Elsevier, Amsterdam.

26. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003.

27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* **6**, 277-293.

28. Johnson, B. A. & Blevins, R. A. (1994). NMR View - a Computer-Program for the Visualization and Analysis of NMR Data. *J Biomol NMR* **4**, 603-614.

29. Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* **104**, 59-107.

30. Schwarzinger, S., Wright, P. E. & Dyson, H. J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry* **41**, 12681-12686.

31. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J. & Dobson, C. M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**, 8977-8991.

32. Rose, G. D. & Roy, S. (1980). Hydrophobic basis of packing in globular proteins. *Proc Natl Acad Sci USA* **77**, 4643-4647.

33. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (Walker, J. M., ed.), pp. 571-607. Humana Press.

34. Wishart, D. S. & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Meth Enzymol* **239**, 363-392.

35. Wang, Y. & Jardetzky, O. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Prot Sci* **11**, 852-861.

36. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **123**, 2970-2978.

37. Plaxco, K. W., Morton, C. J., Grimshaw, S. B., Jones, J. A., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. (1997). The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. *J Biomol NMR* **10**, 221-230.

38. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* **222**, 311-333.

39. Hol, W. G., van Duijnen, P. T. & Berendsen, H. J. (1978). The alpha-helix dipole and the properties of proteins. *Nature* **273**, 443-446.

40. Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* **31**, 1463-1470.

41. Carrington, A. & MacLachlan, A. D. (1969). *Introduction to Magnetic Resonance.*, Harper, New York.

42. Le Duff, C. S., Whittaker, S. B., Radford, S. E. & Moore, G. R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: implications for the early stages of protein folding. *J Mol Biol* **364**, 824-835.

43. Cheng, J., Randall, A. Z., Sweredoski, M. J. & Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.* **33**, W72-W76.

44. Cuff, J. A. & Barton, G. J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* **34**, 508-519.

45. Munoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* **1**, 399-409.

46. Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. & Roder, H. (2001). Structural and kinetic characterization of early folding events in beta-lactoglobulin. *Nat Struct Biol* **8**, 151-155.

47. Reed, M. A., Jelinska, C., Syson, K., Cliff, M. J., Splevins, A., Alizadeh, T., Hounslow, A. M., Staniforth, R. A., Clarke, A. R., Craven, C. J. & Waltho, J. P. (2006). The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J Mol Biol* **357**, 365-372.

48. Fernandez-Recio, J., Genzor, C. G. & Sancho, J. (2001). Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234-15245.

49.   Melo, E. P., Chen, L., Cabral, J. M., Fojan, P., Petersen, S. B. & Otzen, D. E. (2003). Trehalose favors a cutinase compact intermediate off-folding pathway. *Biochemistry* **42**, 7611-7617.

50.   Lopez-Hernandez, E., Cronet, P., Serrano, L. & Munoz, V. (1997). Folding kinetics of Che Y mutants with enhanced native alpha-helix propensities. *J Mol Biol* **266**, 610-620.

51.   Bieri, O. & Kiefhaber, T. (1999). Elementary steps in protein folding. *Biol Chem* **380**, 923-929.

52.   Hiller, S., Wider, G., Imbach, L. L. & Wuthrich, K. (2008). Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX. *Angew Chem Int Ed Engl* **47**, 977-981.

53.   Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.

54.   Le Duff, C. S., Whittaker, S. B. M., Radford, S. E. & Moore, G. R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding. *J Mol Biol* **364**, 824-835.

3

Non-cooperative formation of the off-pathway molten globule during folding of the α-β parallel protein apoflavodoxin

Abstract

During folding of many proteins molten globules are formed. These partially folded forms of proteins have a substantial amount of secondary structure but lack virtually all tertiary side-chain packing characteristic of native structures. Molten globules are ensembles of interconverting conformers and are prone to aggregation, which can have detrimental effects on organisms. Consequently, molten globules attract considerable attention. The molten globule that is observed during folding of flavodoxin from *Azotobacter vinelandii* is a kinetically off-pathway species, as it has to unfold before the native state of the protein can be formed. This intermediate contains helices and can be populated at equilibrium using guanidinium hydrochloride as denaturant, allowing the use of NMR spectroscopy to follow molten globule formation at the residue level. Here, we track changes in chemical shifts of backbone amides, as well as disappearance of resonances of unfolded apoflavodoxin, upon decreasing denaturant concentration. Analysis of the data shows that structure formation within virtually all parts of the unfolded protein precedes folding to the molten globule state. This folding transition is non-cooperative and involves a series of distinct transitions. Four structured elements in unfolded apoflavodoxin transiently interact and subsequently form the ordered core of the molten globule. Although hydrophobic, tryptophan side chains are not involved in the latter process. This ordered core is gradually extended upon decreasing denaturant concentration, but part of apoflavodoxin's molten globule remains random coil in the denaturant range investigated. The results presented here, together with those reported on the molten globule of α-lactalbumin, show that helical molten globules apparently fold in a non-cooperative manner.

Introduction

Folding of proteins to conformations with proper biological activities is of vital importance for all living organisms. To describe folding, the concept of a multi-dimensional energy landscape or folding funnel arose from a combination of experiment, theory and simulation [1-5]. In this model, unfolded protein molecules descend along a funnel describing the free energy of folding, until the folding molecules reach the state that has the lowest free energy, which is the native state. In the energy landscape model, unfolded protein molecules can fold to the native state by following different routes.

Upon folding, proteins can encounter rough folding energy landscapes that allow population of partially folded species, which may be on- or off-pathway to the native state. When the folding species is on-pathway, it is productive for folding. In contrast, when the species is off-pathway, it is trapped as such that the native structure cannot be reached without substantial reorganizational events [6]. A decrease of the folding rate due to the presence of off-pathway intermediates, which are kinetically trapped and partially folded, increases the likelihood of protein aggregation. This aggregation can have detrimental effects on organisms. Folding intermediates are involved during folding of most proteins studied to date [6]. Kinetic intermediates that appear early during folding have been shown to resemble the relatively stable molten globule intermediates, found for several proteins under mildly denaturing conditions [7]. This resemblance suggests that these molten globules can be considered as models of transient intermediates [6].

The term "molten globule" was introduced to describe compact partially folded ensembles of species that have a substantial amount of secondary structure, but lack virtually all tertiary side-chain packing characteristic of native structures [7-10]. The close similarity between molten globule states at equilibrium and those formed during the early stages of refolding has been demonstrated for α-lactalbumin [11-13], apomyoglobin [14, 15], RNAse H [16], T4 lysozyme [17]. Although often poorly defined, molten globule-like intermediates have exposed hydrophobic groups and are prone to forming protein aggregates via specific interactions between structured elements [19]. Inside a living cell ample opportunities exist for aggregation of partially folded protein molecules and this non-productive aggregation can compete with productive folding [20, 21]. Detailed characterization of various equilibrium molten globule states promises to provide valuable insights into the transient kinetic molten globules that are important in folding pathways and in protein aggregation processes.

Here, NMR spectroscopy is used to report the folding at the residue-level of the molten globule-like folding intermediate of flavodoxin, a 179-residue protein from *Azotobacter vinelandii*. Flavodoxins are monomeric proteins involved in electron transport, and contain a non-covalently bound flavin mononucleotide (FMN) cofactor [22]. The proteins consist of a single structural domain and adopt the flavodoxin-like or α-β parallel topology, which is widely prevalent in nature. Both denaturant-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin (i.e., flavodoxin without FMN) have been characterized using guanidine hydrochloride (GuHCl) as denaturant [23-28]. The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to flavodoxin except that residues in the flavin-binding region of the apo protein have considerable dynamics [29, 30]. In presence of FMN, binding of FMN to native apoflavodoxin is the last step in flavodoxin folding.

The kinetic folding of apoflavodoxin involves an energy landscape with two folding intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [25]. Intermediate I_{on} is an obligatory species on the productive route from unfolded to native protein. I_{on} is highly unstable
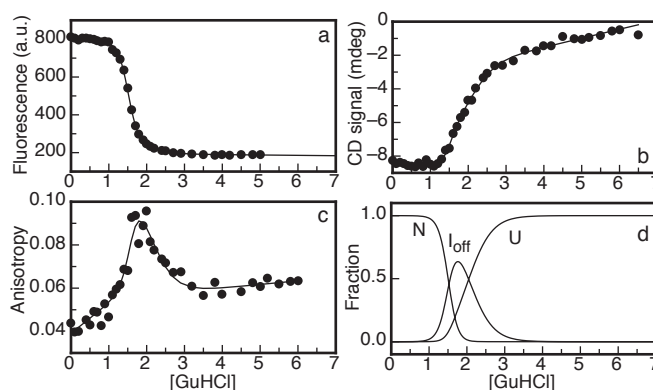
Figure 3.1 Guanidine hydrochloride-induced equilibrium (un)folding of *A. vinelandii* apoflavodoxin [25]. (a) Fluorescence emission at 340 nm with excitation at 280 nm. (b) Circular dichroism at 222 nm. The transition midpoints of the denaturation curves obtained by CD and fluorescence spectroscopy do not coincide, which is characteristic for the population of an intermediate. (c) Fluorescence anisotropy data detected with a 335 nm cut-off filter, excitation is at 300 nm. The curve is biphasic due to population of an intermediate. The solid lines in panels a to c are the result of a global fit of a three-state model for equilibrium (un)folding ($I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin) [25]. (d) Normalized population of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as a function of denaturant concentration. The protein concentration is 5.6 μM in 100 mM potassium pyrophosphate, pH 6.0, and the data are recorded at 25 °C.

and is therefore not observed during denaturant-induced equilibrium unfolding. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [25]. The equilibrium unfolding of apoflavodoxin is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin (Fig. 3.1) [25]. Intermediate I_{off} populates significantly in the concentration range of 1 to 3 M GuHCl (Fig. 3.1d). The off-pathway species is molten globule-like: it is compact, it lacks the characteristic structure of native apoflavodoxin, its three tryptophans are solvent exposed in contrast to the situation in native protein, it contains helices and has severely broadened NMR resonances due to exchange between different conformers on the micro- to millisecond timescale [25, 31, 32]. Elevated protein concentrations [31] and macromolecular crowding [32] cause severe aggregation of this species. The formation of an off-pathway species is typical of proteins with a flavodoxin-like topology [27].

Backbone amide resonances of proteins can be followed in a series of ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra acquired at different denaturant concentrations, as was originally done to study the folding of the molten globule state of α-lactalbumin [33, 34]. Previous experiments, in which this HSQC-based method was used, showed that native apoflavodoxin unfolds highly cooperatively upon addition of denaturant [31] and also indicated that the off-pathway intermediate potentially folds in a non-cooperative manner. However, molecular details of the formation of I_{off} could not be revealed by NMR spectroscopy, because the assignments of NMR resonances of the backbone amides of unfolded apoflavodoxin were lacking.

Recently, heteronuclear NMR spectroscopy has been used to assign resonances of unfolded apoflavodoxin and to characterize the conformational and dynamic properties of the unfolded protein at the residue level [35]. The study shows that the unfolded protein at 6.0 M GuHCl behaves as

a random coil. However, at 3.4 M denaturant, unfolded apoflavodoxin has four transiently ordered regions. These regions have restricted flexibility on the (sub)nanosecond timescale. Three of these regions transiently form α-helices and the fourth ordered region transiently adopts non-native structure, which is neither α-helix nor β-strand [35]. These ordered segments transiently dock non-natively, causing non-productive folding toward the off-pathway folding intermediate I_{off}.

Because NMR assignments of GuHCl-unfolded apoflavodoxin recently became available [35], the above mentioned HSQC-based method can now be used to follow formation of I_{off} at the residue level, as is reported here. At denaturant concentrations that range from 4.05 to 1.58 M GuHCl, 18 ^{1}H-^{15}N HSQC spectra of unfolded apoflavodoxin were acquired. Subsequently, the chemical shifts of the amides of the unfolded protein were determined in these spectra. In addition, the denaturant-dependent disappearance of the amide cross peaks of unfolded apoflavodoxin were tracked. Analysis of the results obtained shows that formation of the molten globule off-pathway species occurs non-cooperatively and is preceded by formation of structured regions in unfolded apoflavodoxin.

Materials and Methods

Protein expression, purification and sample preparation. The single cysteine at position 69 in wild-type *A. vinelandii* (strain ATCC 478) flavodoxin II was replaced by an alanine (Cys69Ala) to avoid covalent dimerisation of apoflavodoxin. This protein variant is largely similar to wild-type flavodoxin regarding both redox potential of holoprotein and stability of apoprotein [23, 36]. Uniformly ^{13}C-^{15}N labeled flavodoxin was obtained from transformed *E. coli* cells grown on ^{13}C-^{15}N labeled algae medium (Silantes, Germany), and purified as described [23].

Unfolded apoflavodoxin was obtained by denaturing flavodoxin in 6 M guanidine hydrochloride (GuHCl, ultrapure from Sigma). Subsequently, FMN was removed via gel filtration at 7 M GuHCl. Two stock solutions of 100 μM protein were prepared in 10% D_2O with 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as internal chemical shift reference. GuHCl concentrations of the stock solutions were 0.06 M and 4.05 M GuHCl, respectively. Subsequently, 550 μl NMR samples of apoflavodoxin at various denaturant concentrations were prepared by appropriately mixing both stock solutions. In total, 18 apoflavodoxin samples with GuHCl concentrations ranging from 4.05 M to 1.58 M GuHCl were prepared. After NMR experiments were finished, refractometry was used to determine the GuHCl concentration in each individual sample [37]. The buffer used in all experiments was 100 mM potassium pyrophosphate (Sigma), pH 6.0.

NMR spectroscopy. Spectra were recorded on a Bruker Avance 700 MHz machine equipped with a triple-resonance 5 mm inverse probe with a self-shielded z-gradient. Sample temperature was 25 ºC. Gradient-enhanced ^{1}H-^{15}N HSQC spectra [38, 39] were recorded of apoflavodoxin samples at different GuHCl concentrations. Spectra were acquired at 4.05, 3.61, 3.41, 3.31, 3.08, 2.94, 2.76, 2.65, 2.55, 2.49, 2.39, 2.25, 2.16, 2.02, 1.92, 1.79, 1.68, or 1.58 M GuHCl. In the ^{1}H dimension of the HSQC experiments, 1024 complex data points were acquired, whereas in the indirect ^{15}N dimension 512 complex data points were collected. Spectral widths were 6010 and 1750 Hz in t_2 and t_1, respectively. With the number of scans set to 16, each HSQC experiment lasted 2 hours and 45 minutes.

To assign the proton resonances of the indole N$^\epsilon$H groups of the three tryptophans of unfolded apoflavodoxin, use was made of three tryptophan variants of apoflavodoxin. These variants

contained a single tryptophan residue (Trp74), or either the Trp74/Trp128 or Trp74/Trp167 pair, which were generated through replacement of tryptophans by phenylalanines using site-directed mutagenesis [40]. One-dimensional proton NMR spectra were recorded of all three tryptophan variants of apoflavodoxin unfolded in 5 M GuHCl.

Determination of cross peak volumes. All spectra were processed with NMRPipe [41] and analyzed with NMRviewJ [42]. Data were zero-filled to 2048 points in both dimensions to gain resolution and smooth the shape of the peaks. The volume of a HSQC cross peak was calculated by summing the intensities of all points within an elliptical region that has its center at the maximum of the cross peak. The line widths at the cross peak's half-height in the ^{1}H and the ^{15}N dimension, respectively, were multiplied by a factor of 1.37 to obtain the values of the radii that define the elliptical region used [43].

Whereas adequate sample shimming is easily achieved, optimal tuning and matching of the probe of the 700 MHz instrument was not possible due to the effects of GuHCl present in the samples. Upon increasing denaturant concentration, the accompanying reduction in NMR receiver coil quality factor Q leads to additional loss of intensity of all signals [31]. A proper correction for these losses is achieved by using HSQC cross peak volumes of backbone amides of residues that behave random coil-like in the range of 4.05 M to 1.58 M GuHCl. The denaturant-dependencies of these cross peak volumes showed no folding transitions due to formation of intermediate I_{off}. Instead, the corresponding chemical shifts change nearly linearly in this denaturant concentration range. Nine residues of unfolded apoflavodoxin showing this behavior were selected (i.e., Arg15, Lys16, Ala18, Lys22, Lys23, Arg24, Glu28, Thr29, and Ser31). The cross peak volumes of the corresponding backbone amides are directly related to the number of nonnative apoflavodoxin molecules (i.e., unfolded and I_{off}) present in the NMR sample. At each denaturant concentration j these nine cross peak volumes were averaged, and the resulting averaged cross peak volume $<V>_j$ was used to properly correct the data for the discussed NMR signal losses. The backbone amide cross peak volume of residue i at denaturant concentration j, i.e., $V_{i,j}$ was divided by $<V>_j$ to obtain the corrected cross peak volume according to:

$$V_{i,j}^{corr} = \frac{V_{i,j}(1 - f_{N,j})}{<V>_j}$$
[3.1]

in which $f_{N,j}$ is the fraction of apoflavodoxin molecules that is native at GuHCl concentration j. The $(1 - f_{N,j})$ term is required to correct for the presence of native apoflavodoxin at GuHCl concentrations below 2 M. These native protein molecules do not contribute to the NMR signals of the unfolded state and cause a further decrease of these signals upon lowering denaturant concentration. Fraction $f_{N,j}$ (Fig. 3.1) is calculated using GuHCl-induced equilibrium (un)folding data of apoflavodoxin [25].

Determination of midpoints of unfolding at the residue level. The disappearance of each individual HSQC cross peak of unfolded apoflavodoxin upon decreasing denaturant concentration could be analyzed according to a two-state model of (un)folding. When a residue is highly flexible, as is the case in an unfolded protein, it gives rise to a very sharp cross peak in a specific region of the HSQC spectrum [35], making this analysis possible. Upon folding, this particular cross peak disappears as backbone motions become restricted.

The free energy of (un)folding is assumed to be linearly dependent on denaturant concentration [44], according to:

$$\Delta G_{U\text{-}F} = \Delta G^{\circ}_{U\text{-}F} + m[D] \qquad\qquad [3.2]$$

with $\Delta G_{U\text{-}F}$ the difference in free energy between the unfolded state and a folded state at a particular denaturant concentration. $\Delta G^{\circ}_{U\text{-}F}$ is $\Delta G_{U\text{-}F}$ at zero denaturant concentration, m the slope at the midpoint of the unfolding transition, and $[D]$ the concentration of denaturant, respectively. It follows from Equation 3.2 that at C_m [45]:

$$\Delta G^{\circ}_{U\text{-}F} = \text{-}m\cdot C_m \qquad\qquad [3.3]$$

Thus:

$$\Delta G_{U\text{-}F} = m([D]\text{-}C_m) \qquad\qquad [3.4]$$

Incorporation of Equation 3.4 into the linear extrapolation method [46], which takes into account the linear dependence of the pre- and post-unfolding data on the denaturant concentration, leads to the following general equation:

$$Y_{obs} = \frac{(\alpha_F+\beta_F[D])+(\alpha_U+\beta_U[D])\cdot e^{\frac{-(m[D]\text{-}C_m)}{RT}}}{1 + e^{\frac{-(m[D]\text{-}C_m)}{RT}}} \qquad\qquad [3.5]$$

with Y_{obs} representing any physical parameter that characterizes the folded and unfolded states of a protein at a particular denaturant concentration, α and β are intercepts and slopes of pre- and post- unfolding regions, and RT is 0.59 kcal mol^{-1}.

In case of the apoflavodoxin folding data, obtained by a series of denaturant-dependent HSQC spectra, cross peaks that solely correspond with highly flexible residues were followed. As there is no interference of these cross peaks with those arising from residues that have restricted motions due to folding, α_F and β_F can be set to 0 in Equation 3.5, resulting in:

$$Y_{obs} = \frac{(\alpha_U+\beta_U[D])\cdot e^{\frac{-(m[D]\text{-}C_m)}{RT}}}{1 + e^{\frac{-(m[D]\text{-}C_m)}{RT}}} \qquad\qquad [3.6]$$

The residue-specific C_m values were calculated by applying Equation 3.6 to the denaturant-dependent corrected cross peak volumes of 68 residues of unfolded apoflavodoxin. If a cross peak was no longer observed above noise level, the corresponding cross peak volume was set to zero. Cross peak volumes at 0, 0.3, 0.6 and 0.9 M GuHCl of these 68 residues were set to zero, as no highly flexible residues typical for unfolded protein were experimentally detected at these denaturant concentrations [31]. Indeed, the GuHCl-induced equilibrium (un)folding data of apoflavodoxin show that the protein is native up to 0.9 M GuHCl [25].

Fitting of denaturant-dependent chemical shift changes. Chemical shifts of 114 backbone amides of unfolded apoflavodoxin could be followed in the 18 HSQC spectra obtained in the denaturant range of 4.05 M to 1.58 M GuHCl. Depending on how ^{1}H and ^{15}N chemical shifts change upon altering denaturant concentration, either a linear, exponential or polynomial function was used to fit to these data. Subsequently, for each NH group at a particular denaturant concentration [D], the magnitude of the non-linear dependence of chemical shifts on denaturant concentration ($X_{[D]}$) was calculated according to:

$$X_{[D]} = \left| \frac{d^2(\delta^1 H)}{d^2[D]} \right| + 10 \cdot \left| \frac{d^2(\delta^{15} N)}{d^2[D]} \right| \tag{3.7}$$

in which δ^1H is the proton chemical shift and δ^{15}N the nitrogen chemical shift. When $X_{[D]}$ is above zero, the chemical shift of either ^{1}H or ^{15}N, or of both nuclei, has a non-linear dependency on the denaturant concentration.

Results and Discussion

Residue-specific folding of apoflavodoxin's molten globule can be followed by NMR spectroscopy. Gradient-enhanced ^{1}H-^{15}N HSQC spectra were recorded of 18 apoflavodoxin samples with GuHCl concentrations ranging from 4.05 to 1.58 M GuHCl (Fig. 3.2). In this denaturant range, the transition of unfolded protein to I_{off} should be detectable (Fig. 3.1d). Due to the relatively low protein concentration used, i.e., 100 µM, no protein aggregation is observed in these NMR samples. In the HSQC spectra, cross peaks of unfolded apoflavodoxin were assigned using the assignments of all H^N and ^{15}N resonances of unfolded apoflavodoxin in 6.0 M and in 3.4 M GuHCl [35] and by using a 3D HNCA spectrum of apoflavodoxin in 2.22 M GuHCl.

In the denaturant range from 4.0 to 1.6 M GuHCl, exchange between unfolded apoflavodoxin and I_{off} is slow on the NMR chemical shift timescale (i.e., ~ 10 s^{-1} at 4.0 M GuHCl and ~ 60 s^{-1} at 1.6 M GuHCl) [25]. Consequently, upon decreasing the denaturant concentration, cross peaks arising from unfolded apoflavodoxin should disappear from the corresponding HSQC spectra, and cross peaks arising from I_{off} should appear. However, as I_{off} is a molten globule, most of its resonances are not observed because these resonances are broadened beyond detection [31]. This severe line broadening is due to conformational exchange processes on the micro- to millisecond timescale between the ensemble of conformers that represents the molten globule [10].

At 1.58 M GuHCl, the lowest denaturant concentration used, 34% of the total number of protein molecules is native (Fig. 3.1d) [25]. At this denaturant concentration, conformational exchange between native and unfolded apoflavodoxin is slow on the NMR chemical shift time scale [31]. However, due to the low protein concentration used, at 1.58 M GuHCl no backbone amide cross peaks are observed for native apoflavodoxin in the corresponding HSQC spectrum (except for the highly flexible C-terminal Leu179). In contrast, cross peaks arising from unfolded apoflavodoxin, which has large intrinsic flexibility [35], are much sharper and thus visible.

In conclusion, upon decreasing GuHCl concentration from 4.05 M to 1.58 M, HSQC spectra show almost exclusively cross peaks arising from unfolded apoflavodoxin. Tracking the denaturant-dependent changes in the corresponding cross peak volumes and chemical shifts enables the study of the residue-specific folding of the unfolded protein to I_{off}.
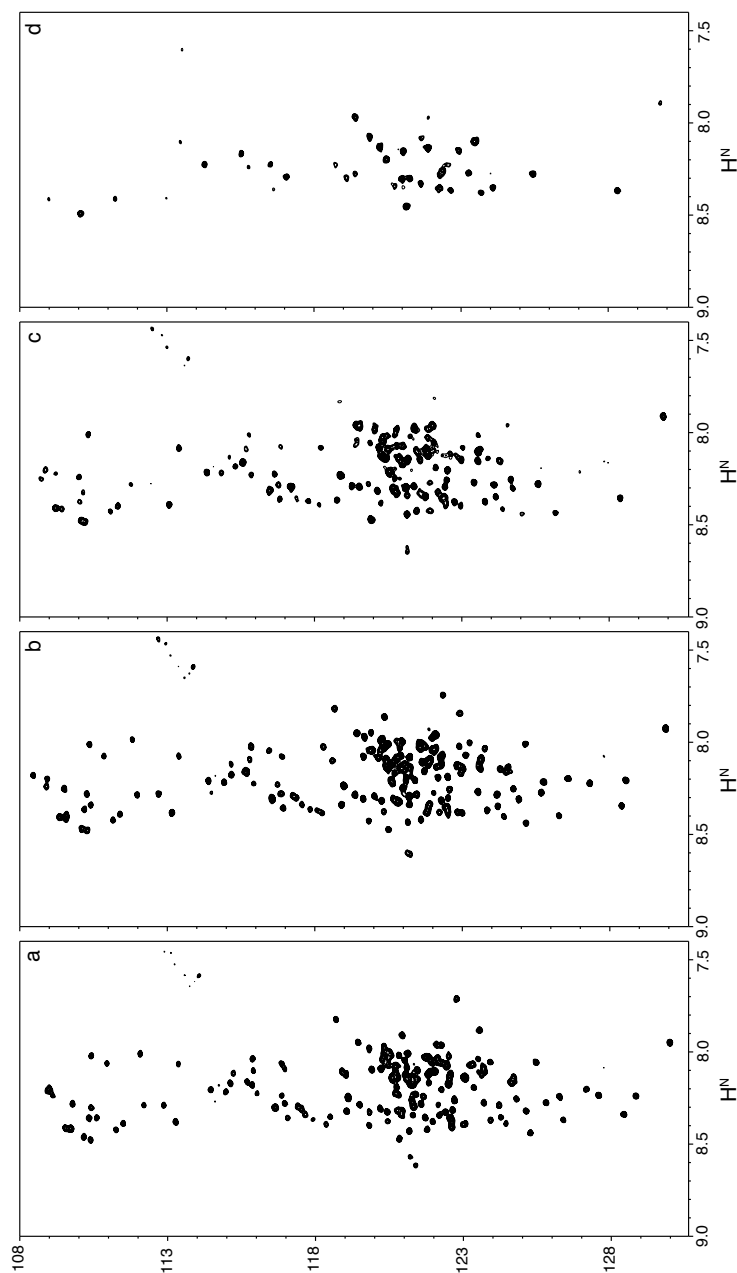
Figure 3.2 NMR resonances of unfolded apoflavodoxin disappear upon folding of the protein. Shown are ^{1}H-^{15}N HSQC spectra of apoflavodoxin in (a) 4.05 M GuHCl, (b) 3.08 M GuHCl, (c) 2.39 M GuHCl, and (d) 1.58 M GuHCl. Protein concentration is 100 μM in 100 mM KPPi, pH 6.0, 25 °C.

Several residues have comparable dynamical and conformational properties in I_{off} and unfolded apoflavodoxin. Figure 3.2 shows HSQC spectra of apoflavodoxin in different GuHCl concentrations. The HSQC spectrum obtained at 4.05 M GuHCl (Fig. 3.2a) is typical for an unfolded protein and has limited dispersion in the ¹H dimension, whereas dispersion in the ¹⁵N dimension is good. All cross peaks observed at 4.05, 3.61, 3.41, 3.31, 3.08 and 2.94 M GuHCl are sharp, as expected for an unfolded protein. Upon decreasing the GuHCl concentration from 4.05 M to 2.94 M, the number of cross peaks detected remains constant and several cross peaks shift. Upon decreasing the denaturant concentration to 2.76 M GuHCl, the volume of some cross peaks reduces. The latter phenomenon becomes more pronounced upon further decreasing denaturant concentration, and at 2.48 M GuHCl a few cross peaks have disappeared from the HSQC spectrum. At 1.58 M GuHCl, many more cross peaks are not observed in the corresponding HSQC spectrum (Fig. 3.2d).

The majority of backbone amide cross peaks of unfolded apoflavodoxin disappears upon decreasing the denaturant concentration from 4.05 M to 1.58 M GuHCl. However, 27 sharp cross peaks with considerable intensity are still present at 1.58 M GuHCl (Fig. 3.2d) and arise from the following residues: Phe6, Lys13 - Lys16, Ala18, Lys19, Ile21 - Asp32, Leu34, Val36, Gly60, Asp68 - Glu70, and Ser178. These 27 cross peaks exhibit only small chemical shift changes upon
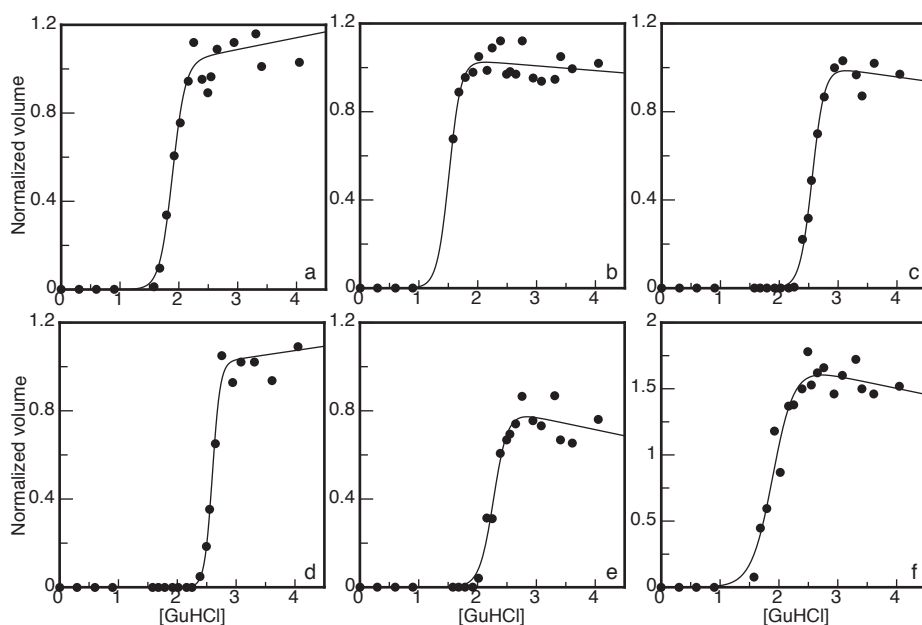
Figure 3.3 Folding of unfolded apoflavodoxin at the residue-level, as monitored by changes in cross peak volumes in ¹H-¹⁵N HSQC spectra at various concentrations of GuHCl. Shown are cross peak volumes of the backbone amides of (a) Gly4, (b) Ala18, (c) Gly53, (d) Ala92, (e) Leu179, and of (f) the indole NεH group of Trp128. All cross peaks are corrected for GuHCl-induced changes in volume that are not due to protein folding, as described in Materials and Methods. Cross peak volumes at 0, 0.3, 0.6 and 0.9 M GuHCl are set to zero, as no cross peaks typical for unfolded protein were detected at these denaturant concentrations. The solid curves show the fit of a two-state model of protein folding (i.e., Equation 3.6, Materials and Methods) to the apoflavodoxin folding data. The extracted midpoints of folding (i.e., C_m) are: 1.89 M (Gly4), 1.51 M (Ala18), 2.56 M (Gly53), 2.60 M (Ala92), 2.28 M (Leu179), and 1.90 M (NεH Trp128).

decreasing the denaturant concentration. Equilibrium unfolding studies show that at 1.58 M GuHCl, 34% of the protein molecules is native, 53% is off-pathway folding intermediate and only 13% is unfolded (Fig. 3.1) [25]. Consequently, at 1.58 M GuHCl, the 27 residues mentioned must have comparable dynamical and conformational properties in I_{off} and unfolded apoflavodoxin.

Unfolded apoflavodoxin folds non-cooperatively to I_{off}. To study the residue-specific folding of unfolded apoflavodoxin to I_{off}, denaturant-dependent changes in the ^{1}H-^{15}N cross peak volumes of unfolded protein are followed. Cross peak volumes instead of cross peak intensities are used to correct for changes in R_2 relaxation times, as several regions of unfolded apoflavodoxin become transiently ordered [35], and as upon reducing GuHCl concentration the viscosity of the solution diminishes. Due to the presence of GuHCl in the apoflavodoxin samples, intensities of all NMR signals differ slightly between the different HSQC spectra [31]. A proper correction for these intensity changes, which are not due to protein folding, is possible and is applied to all cross peak volumes, as described in Materials and Methods.

Changes in corrected cross peak volumes of 68 well-resolved backbone amides and of the 3 tryptophan indole $N^{\epsilon}H$ groups of unfolded apoflavodoxin could be followed upon decreasing the denaturant concentration. Some examples of folding transitions observed are shown in Figure 3.3. A two-state model suffices to describe each folding transition.

The apparent midpoints of folding (i.e., C_m values, see Materials and Methods) of the backbone amides of 68 residues of unfolded apoflavodoxin are shown in Figure 3.4 and listed in Supporting Information, Table S3.1 (including C_m values of the tryptophan indole $N^{\epsilon}H$ groups). The residues investigated are distributed along the entire sequence of apoflavodoxin. Significant differences exist between C_m values, since they range from 1.5 M to 2.7 M GuHCl, and five clusters of residues can be identified (Fig. 3.4b), although the boundaries of these clusters are not sharply defined.

The slope associated with the folding transition from unfolded apoflavodoxin to I_{off} as observed by far-UV CD is rather shallow (Fig. 3.1b). In contrast, each folding transition detected at the residue level by NMR spectroscopy has a much steeper slope (Fig. 3.3). The residue-specific C_m values cover a broad range of denaturant concentrations. Folding toward I_{off} involves a series of distinct transitions and as a result far-UV CD reports a shallow folding transition. The HSQC data presented here show that folding of I_{off} occurs non-cooperatively.

Upon decreasing denaturant concentration, the ordered core of I_{off} becomes gradually extended. At 3.4 M GuHCl four transiently ordered regions are detected in unfolded apoflavodoxin (comprising residues 41 – 53, 72 – 83, 99 – 122 and 160 – 176). These regions transiently dock onto one another in a non-native manner, and this non-native docking promotes folding toward the helical off-pathway intermediate [35].

The residue specific folding data presented here show that the residues that have the largest average C_m value (i.e., Group-1, $C_m^{average} = 2.64 \pm 0.05$ M GuHCl; red in Figure 3.4; typical examples Gly53 and Ala92, Figures 3.3c and d) roughly coincides with those residues that are transiently ordered in unfolded apoflavodoxin (gray bars in Figure 3.4a). Thus, upon decreasing denaturant concentration, assembly of these four structured elements in unfolded apoflavodoxin leads to formation of the ordered core of I_{off}.

The C_m data of the group of residues that has the second largest average C_m value (i.e., Group-2, $C_m^{average} = 2.40 \pm 0.07$ M GuHCl) are highlighted in yellow in Figure 3.4. All corresponding residues are close in sequence to the residues of Group-1, which together form the core of I_{off}. This
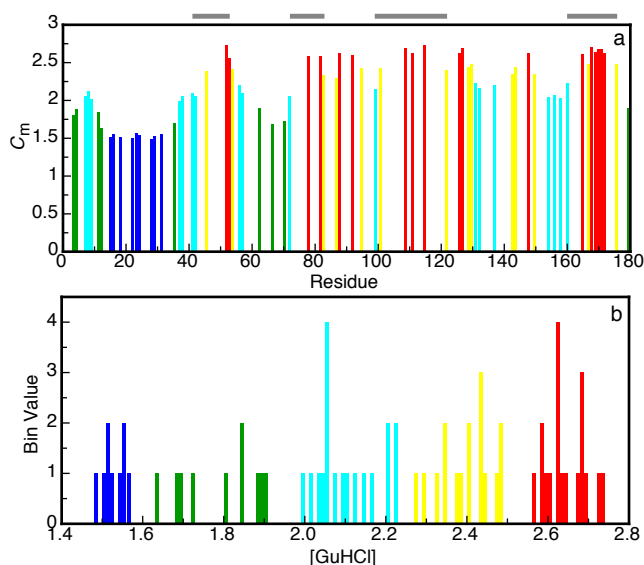
Figure 3.4 Midpoints of folding (C_m values) determined for the backbone amide groups and the three indole NεH groups of unfolded apoflavodoxin. (a) C_m values are plotted versus residue number, errors range between 0.002 and 0.075 M; each vertical bar is colored according to the coloring scheme of (b). Horizontal gray bars highlight the four transiently ordered regions in unfolded apoflavodoxin at 3.4 M GuHCl [35]. (b) Histogram of binned C_m values; the bin width is 0.01 M GuHCl. Five groups of residues are identified and are colored differently. Group-1 (C_m between 2.5 M to 2.8 M GuHCl): red, Group-2 (C_m between 2.2 M to 2.5 M GuHCl): yellow, Group-3 (C_m between 1.95 M and 2.2 M GuHCl): cyan, Group-4 (C_m between 1.6 M and 1.95 M GuHCl): green, and Group-5 (C_m between 1.4 M and 1.6 M GuHCl): dark blue.

observation suggests that upon decreasing denaturant concentration the initial ordered core of I_{off} becomes extended.

A further reduction in denaturant concentration leads to additional sequential extension of the folded regions of I_{off}. This conclusion is drawn on basis of the C_m data of the group of residues that has the third largest average C_m value (i.e., Group-3, $C_m^{average}$ = 2.10 ± 0.07 M GuHCl, cyan in Figure 3.4). This distribution shows that most of the corresponding residues are close in sequence to those residues that already became ordered in I_{off} at higher denaturant concentrations. However, residues 7 to 9 cooperatively fold to newly formed structure in I_{off} and their folding does not extend existing ordered parts of I_{off}.

The fourth group of residues of unfolded apoflavodoxin has more widely distributed C_m values (i.e., Group-4, $C_m^{average}$ = 1.79 ± 0.10 M GuHCl, green in Figure 3.4; typical example Leu179, Figure 3.3e). Most of these residues reside in the N-terminal part of the protein and are positioned at the N- and C-terminal sides of ordered residues 7 to 9, thereby extending the structure formed by the latter residues.

Finally, NMR spectroscopy shows that the remaining fifth group of residues of unfolded apoflavodoxin has an average C_m value of 1.53 ± 0.03 M GuHCl (i.e., Group-5, dark blue in Figure 3.4; typical example Ala18, Figure 3.3b). This $C_m^{average}$ is within error identical to the C_m values derived from denaturant-induced unfolding of native apoflavodoxin detected by fluorescence emission spectroscopy (C_m = 1.53 ± 0.01 M GuHCl, Fig. 3.1a [25]) and by NMR spectroscopy

(C_m = 1.48 ± 0.04 M GuHCl [31]). Thus, the C_m values obtained for Arg15, Lys16, Ala18, Lys22, Lys23, Arg24, Glu28, Thr29, and Ser31 report folding of unfolded apoflavodoxin to native protein. Consequently, in the denaturant range investigated, these residues are unfolded in I_{off}.

Tryptophans do not cause folding of unfolded apoflavodoxin to I_{off}. Apoflavodoxin contains three tryptophans (i.e., Trp74, Trp128 and Trp167). Of these tryptophans, both Trp74 and Trp167 belong to regions with restricted flexibility in unfolded apoflavodoxin in 3.4 M GuHCl, as revealed by elevated ^{15}N R_2 relaxation rates of the corresponding backbone amides [35]. Analysis of the denaturant-dependent disappearance of the HSQC cross peaks arising from the indole N$^\epsilon$H groups of these tryptophans of unfolded apoflavodoxin gives the following C_m values: Trp74-N$^\epsilon$H: 1.84 M, Trp128-N$^\epsilon$H: 2.28 M, and Trp167-N$^\epsilon$H: 2.38 M (Table S3.1). These values are significantly smaller than the average C_m value of 2.64 ± 0.05 M that characterizes the folding of the residues of Group-1 (Fig. 3.4).

As discussed, docking of the transiently ordered residues of Group-1 in unfolded apoflavodoxin leads to formation of the ordered core of I_{off}. The relatively low C_m values obtained for the N$^\epsilon$H groups of the tryptophans of unfolded apoflavodoxin imply that the side chains of these amino acid residues are not involved in this process. Apoflavodoxin's tryptophans do not have substantial hydrophobic interactions with the ordered regions of the unfolded protein. In contrast, in case of lysozyme the hydrophobic tryptophan at position 62 plays an important role in folding, as it is involved in extensive long-range tertiary interactions in the urea-denatured state of the protein [47]. Apparently, tryptophans do not play a significant role in folding of I_{off}.

Transient formation of ordered structure within virtually all parts of unfolded apoflavodoxin precedes transition to I_{off}. To further clarify the folding events that occur upon decreasing the GuHCl concentration, ^{1}H and ^{15}N chemical shifts of unfolded apoflavodoxin are followed. Changes in chemical shifts are good indicators for formation of ordered structure in unfolded proteins [35, 48]. Chemical shifts of cross peaks of 114 backbone amides of unfolded apoflavodoxin have been followed in the 18 HSQC spectra discussed. These cross peaks were selected because they do not severely overlap with each other.

Chemical shifts of backbone amides that depend linearly on denaturant concentration (Fig. 3.5a,b) merely reflect the change in average properties of the solvent surrounding these amides. Thus, no structure formation involving the corresponding residues takes place in unfolded apoflavodoxin. In contrast, chemical shifts of various backbone amides have a non-linear dependence on denaturant concentration (Fig. 3.5c-f), reflecting transient structure formation in unfolded apoflavodoxin molecules.

Depending on how the ^{1}H and ^{15}N chemical shifts change as a function of denaturant concentration, either a linear, exponential or polynomial function was used to fit the data. Subsequently, at a particular denaturant concentration [D] the magnitude of the non-linear dependence of the chemical shift on denaturant concentration (i.e., $X_{[D]}$) is calculated per residue by taking the second derivatives of the corresponding fits (see Materials and Methods). $X_{[D]}$ is used as a measure of the increase in structure formation at the residue level in unfolded apoflavodoxin upon decreasing denaturant concentration. $X_{[D]}$ data are presented in Figure 3.6.

Figure 3.6 shows that residues 21 to 35 of unfolded apoflavodoxin are characterized by $X_{[D]}$-values of almost zero at all denaturant concentrations used. Thus, no structure is formed in this region of unfolded apoflavodoxin. Note that these residues show no transition from unfolded apoflavodoxin
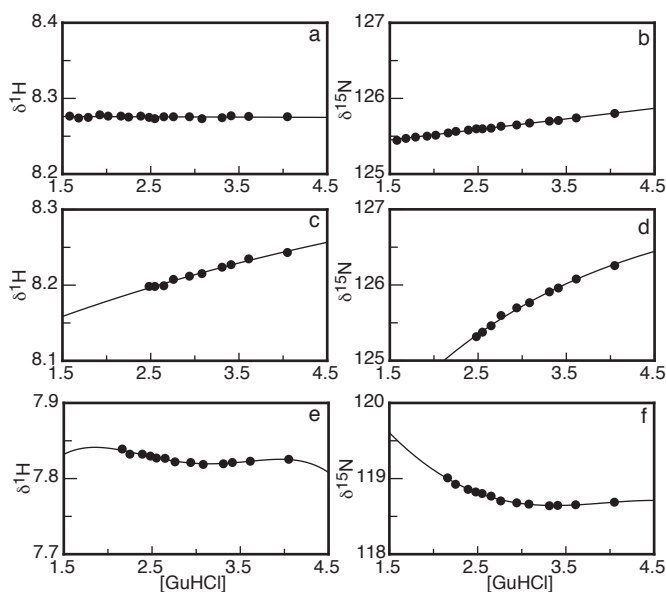
Figure 3.5 Denaturant-dependence of chemical shifts of various backbone amides of unfolded apoflavodoxin. Chemical shifts are shown for (a) ^{1}H and (b) ^{15}N of the backbone amide of Lys22, (c) ^{1}H and (d) ^{15}N of the backbone amide of Leu52, and (e) ^{1}H and (f) ^{15}N of the backbone amide of Trp167. Lines show the fit of a linear (a, b), exponential (c, d) or polynomial function (e, f) to the data.

to I_{off} in the denaturant range investigated. However, for all other residues, $X_{[D]}$ is well above zero at denaturant concentrations close to their C_m-values. Consequently, these residues form transient structure in unfolded protein. Figure 3.6 shows that the first regions that become ordered upon decreasing denaturant concentration are the four regions that are transiently ordered in unfolded apoflavodoxin [35]. Non-native docking of these regions leads to formation of I_{off}. In addition, Figure 3.6 shows that upon decreasing GuHCl concentration, most of the remaining parts of the unfolded protein also become transiently ordered.

Clearly, virtually all residues of unfolded apoflavodoxin form transient structure upon decreasing denaturant concentration, and this process precedes the transition of the unfolded species to the off-pathway folding intermediate I_{off}.

Part of the off-pathway folding intermediate I_{off} is random coil. As discussed, at 1.58 M GuHCl, the lowest denaturant concentration used in this study, 27 residues of apoflavodoxin have the dynamical and conformational properties observed for an unfolded protein. Most of these residues are clustered in two regions of apoflavodoxin; Lys13 to Val36 (Region-U1), and Gly60 to Glu70 (Region-U2). Determination of the denaturant-dependent changes in chemical shifts of the corresponding backbone amides (Fig. 3.6), as well as the disappearance of the corresponding resonances (Fig. 3.4a), enables the further characterization of these two regions.

Analysis of C_m values shows that at 1.58 M GuHCl Region-U1 must be unfolded in the ensemble of species that forms I_{off}, as no folding transition from unfolded protein to I_{off} is observed (Fig. 3.4a; Group-5, dark blue). Figure 3.6 shows that in the range from 4.05 M to 1.58 M GuHCl, part of Region-U1 (i.e., residues 21 to 36) has $X_{[D]}$ values close to or equal to zero. Thus, in this denaturant
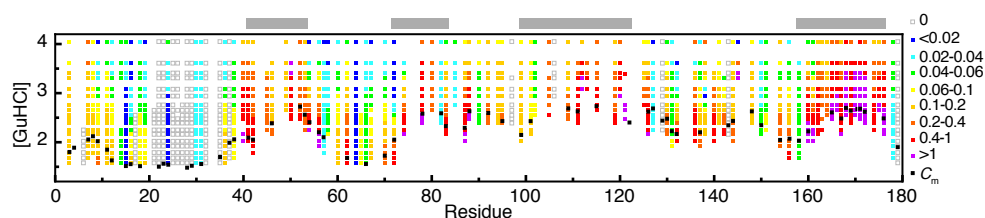
Figure 3.6 Magnitude $X_{[D]}$ of the non-linear dependence of the ^1H and ^{15}N chemical shifts of unfolded apoflavodoxin on denaturant concentration versus residue number. The vertical axis shows the GuHCl concentration, and coloration indicates the magnitude of $X_{[D]}$ (see Materials and Methods). Horizontal gray bars highlight the four transiently ordered regions in unfolded apoflavodoxin [35]. The midpoints of unfolding determined for the backbone amide groups of unfolded apoflavodoxin (Fig. 3.4a) are indicated by (-).

range no structure formation in the C-terminal part of Region-U1 takes place. Consequently, at 1.58 M GuHCl, the C-terminal part of Region-U1 behaves as a random coil. Whether this random coil behavior persists upon further decreasing denaturant concentration is unknown. The N-terminal part of Region-U1 has relatively small $X_{[D]}$ values in the denaturant range investigated. Consequently, at 1.58 M GuHCl, the N-terminal part of Region-U1 is involved in some minor structure formation. Upon further decreasing denaturant concentration, this protein part likely becomes structured in I_{off}.

The backbone amides of Region-U2 of apoflavodoxin are characterized by an average C_m value of 1.77 M GuHCl. The corresponding residues thus report the folding transition that links unfolded apoflavodoxin with I_{off}. Therefore, these residues must be ordered in I_{off}. At 1.58 M GuHCl, which is still in the transition region of denaturant-induced equilibrium (un)folding of these residues, Region-U2 is unfolded in part of the ensemble of apoflavodoxin molecules. Consequently, at 1.58 M GuHCl, the corresponding backbone amide cross peaks are still observable in the HSQC spectrum. Upon further decreasing denaturant concentration, and in absence of denaturant, all residues of Region-U2 of apoflavodoxin will become ordered in I_{off}.

Molten globules of helical nature apparently fold in a non-cooperative manner. In unfolded apoflavodoxin several regions of the protein have reduced flexibility, which is due to transient helix formation and local and non-local hydrophobic interactions. Helices are the only regular secondary structure elements detected in unfolded apoflavodoxin and these helices are sufficiently stable to be present about 10% of the time [35].

Upon protein folding, helices are formed much more rapidly than sheets, especially when parallel β-sheets are involved. This rapid helix formation is due to the highly local character of the interactions in helices, whereas the residues that have to be brought into contact to form a parallel β-sheet, as is the case for apoflavodoxin, are separated by many residues from one another [49, 50].

Rapid formation of three α-helices and transiently ordered structure that is neither α-helix nor β-strand and their subsequent non-native docking through hydrophobic interactions causes formation of the off-pathway intermediate during apoflavodoxin folding [35]. This docking of helices prevents formation of the parallel β-sheet of apoflavodoxin and causes the intermediate to be helical. The helical character of the intermediate is confirmed by far-UV CD measurements (see chapter 5 of this thesis). Similar processes are expected to play a role during kinetic folding of other proteins with an α-β parallel topology [35], causing them to be susceptible to off-pathway intermediate formation. Indeed, an off-pathway intermediate is experimentally observed for all other α-β parallel proteins of

which the kinetic folding has been investigated: apoflavodoxin from *Anabaena* [51], CheY [52], cutinase [53], and UMP/CMP kinase [54].

Whereas previous NMR experiments showed that formation of native apoflavodoxin is highly cooperative [31], the data presented here show that I_{off} is non-cooperatively formed. This non-cooperativity is probably due to the helical character of the off-pathway intermediate. Helices involve relatively short-range interactions and the folding of one helix may not affect the folding of other helices [55]. Indeed, in case of the α-helical domain of the molten globule of α-lactalbumin, which has native-like secondary structure, the α-helices fold at differing denaturant concentrations [33]. In contrast, the folding of the molten globule of human serum retinol-binding protein, which contains a significant amount of β-sheet structure, is significantly more cooperative. The formation of a stable β-sheet requires that at least two β-strands fold and interact, which is the cause of this cooperativity. Our observations concerning the folding of the off-pathway intermediate of apoflavodoxin, together with those of the molten globule of α-lactalbumin, show that the folding of helical molten globules is apparently non-cooperative.

Supporting Information. *Table S3.1* Midpoints of unfolding (C_m values) determined for the backbone amide groups and the three NεH indole groups of unfolded apoflavodoxin. Five groups of residues are identified based on their C_m-values and have been binned.

Residue	C_m (M)	st. error	Bin group
3	1.80202	4.94 e-3	4
4	1.88867	3.41 e-3	4
7	2.05607	2.54 e-3	3
8	2.12187	3.10 e-3	3
9	2.01959	0.01014	3
11	1.84865	4.66 e-3	4
12	1.6304	0.01364	4
15	1.51523	7.59 e-3	5
16	1.5542	4.17 e-3	5
18	1.51308	7.14 e-3	5
22	1.50227	7.35 e-3	5
23	1.56308	3.19 e-3	5
24	1.54275	8.10 e-3	5
28	1.48193	7.40 e-3	5
29	1.521	7.30 e-3	5
31	1.55629	7.01 e-3	5
35	1.6996	5.24 e-3	4
37	1.99031	6.02 e-3	3
38	2.05904	2.85 e-3	3
41	2.09333	3.78 e-3	3
42	2.0529	0.01181	3
46	2.38505	5.15 e-3	2
52	2.72596	3.99 e-3	1
53	2.56062	3.81 e-3	1
54	2.40814	3.73 e-3	2
56	2.20392	1.98 e-3	3
57	2.10216	2.49 e-3	3
62	1.89319	3.22 e-3	4
66	1.6809	0.01253	4
70	1.72647	7.24 e-3	4
72	2.05219	4.53 e-3	3
78	2.58171	3.44 e-3	1
82	2.58923	5.09 e-3	1
83	2.32871	6.04 e-3	2
87	2.29341	7.02 e-3	2
88	2.6312	4.96 e-3	1
92	2.59555	2.56 e-3	1
95	2.43094	7.18 e-3	2
99	2.14808	4.24 e-3	3
101	2.43004	5.61 e-3	2
109	2.6895	4.73 e-3	1
111	2.62842	8.71 e-3	1
115	2.73492	3.33 e-3	1
122	2.40192	3.77 e-3	2
126	2.62371	3.38 e-3	1
127	2.688	9.51 e-3	1
129	2.44036	0.01845	2
130	2.47884	0.04081	2
131	2.22619	5.98 e-3	2
132	2.16724	3.89 e-3	3
137	2.20321	5.33 e-3	3
143	2.34586	7.28 e-3	2
144	2.4344	0.0751	2
148	2.62897	6.36 e-3	1
150	2.34721	3.42 e-3	2
154	2.04744	2.96 e-3	3
156	2.07243	5.11 e-3	3
158	2.03176	0.0101	3
160	2.22182	4.08 e-3	2
165	2.6074	0.04835	1
167	2.48431	3.62 e-3	2
168	2.69964	2.69 e-3	1
169	2.64298	5.06 e-3	1
170	2.67734	0.02761	1
171	2.6825	0.03857	1
172	2.62101	3.26 e-3	1
176	2.48529	9.04 e-3	2
179	1.90008	3.52 e-3	4
NεH 74	1.84349	2.53 e-3	4
NεH 128	2.27556	5.37 e-3	2
NεH 167	2.3756	4.08 e-3	2

References

1. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.
2. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19.
3. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25**, 331-339.
4. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641-645.
5. Fersht, A. R. & Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell* **108**, 573-582.
6. Jahn, T. R. & Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-117.
7. Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209-282.
8. Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Letters* **164**, 21-24.
9. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Adv Protein Chem* **47**, 83-229.
10. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-132.
11. Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold Des* **1**, 275-287.
12. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. & Dobson, C. M. (1999). Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J Mol Biol* **288**, 673-688.
13. Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K., Kihara, H., Amemiya, Y. & Kuwajima, K. (2002). Fast compaction of alpha-lactalbumin during folding studied by stopped-flow X-ray scattering. *J Mol Biol* **321**, 121-132.
14. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-1548.
15. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896.
16. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
17. Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* **365**, 870-880.
18. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-226.
19. Jaenicke, R. & Seckler, R. (1997). Protein misassembly in vitro. *Adv Protein Chem* **50**, 1-59.
20. Ellis, R. J. & Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Curr Opin Struct Biol* **9**, 102-110.
21. Hartl, F. U. & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858.
22. Mayhew, S. G. & Tollin, G. (1992). General properties of flavodoxins. In *Chemistry and biochemistry of flavoenzymes* (Muller, F., ed.), Vol. 3, pp. 389-426. CRC press, Boca Raton, Florida.
23. van Mierlo, C. P. M., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Sci* **7**, 2331-2344.
24. van Mierlo, C. P. M. & Steensma, E. (2000). Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *J Biotechnol* **79**, 281-298.
25. Bollen, Y. J., Sanchez, I. E. & van Mierlo, C. P. M. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry* **43**, 10475-10489.
26. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem* **280**, 7836-7844.

27. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

28. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* **103**, 4095-4100.

29. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of Azotobacter vinelandii holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

30. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

31. van Mierlo, C. P. M., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

32. Engel, R., Westphal, A. H., Huberts, D. H., Nabuurs, S. M., Lindhoud, S., Visser, A. J. & van Mierlo, C. P. M. (2008). Macromolecular Crowding Compacts Unfolded Apoflavodoxin and Causes Severe Aggregation of the Off-pathway Intermediate during Apoflavodoxin Folding. *J Biol Chem* **283**, 27383-27394.

33. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol* **4**, 630-634.

34. Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S. & Dobson, C. M. (1999). Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat Struct Biol* **6**, 948-952.

35. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. (2008). Extensive formation of off-pathway species during folding of an alpha-beta parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J Am Chem Soc* **130**, 16914-16920

36. Steensma, E., Heering, H. A., Hagen, W. R. & van Mierlo, C. P. M. (1996). Redox properties of wild-type, Cys69Ala, and Cys69Ser Azotobacter vinelandii flavodoxin II as measured by cyclic voltammetry and EPR spectroscopy. *Eur J Biochem* **235**, 167-172.

37. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. In *Methods in enzymology* (Hirs, C. H. W. & Timasheff, S. N., eds.), Vol. 26, pp. 43-50. Academic Press, New York.

38. Palmer, A. G., 3rd, Cavanagh, J., Wright, P. E. & Rance, M. (1991). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **93**, 151-170.

39. Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665.

40. Visser, N. V., Westphal, A. H., van Hoek, A., van Mierlo, C. P. M., Visser, A. J. W. G. & van Amerongen, H. (2008). Tryptophan-tryptophan energy migration as a tool to follow apoflavodoxin folding. *Biophys J* **95**, 2462-2469.

41. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* **6**, 277-293.

42. Johnson, B. A. & Blevins, R. A. (1994). NMR View - a Computer-Program for the Visualization and Analysis of NMR Data. *J Biomol NMR* **4**, 603-614.

43. Rischel, C. (1995). Fundamentals of Peak Integration. *J Mag Res Series A* **116**, 255-258.

44. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266-280.

45. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**, 11259-11269.

46. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068.

47. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.

48. Dyson, H. J. & Wright, P. E. (2005). Elucidation of the protein folding landscape by NMR. *Methods Enzymol* **394**, 299-321.

49. Bieri, O. & Kiefhaber, T. (1999). Elementary steps in protein folding. *Biol Chem* **380**, 923-929.

50. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-994.

51. Fernandez-Recio, J., Genzor, C. G. & Sancho, J. (2001). Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234-15245.

52. Kathuria, S. V., Day, I. J., Wallace, L. A. & Matthews, C. R. (2008). Kinetic Traps in the Folding of βα-Repeat Proteins: CheY Initially Misfolds before Accessing the Native Conformation. *J Mol Biol* **382**, 467-484.

53. Otzen, D. E., Giehm, L., Baptista, R. P., Kristensen, S. R., Melo, E. P. & Pedersen, S. (2007). Aggregation as the basis for complex behaviour of cutinase in different denaturants. *Biochim Biophys Acta - Proteins & Proteomics* **1774**, 323-333.

54. Lorenz, T. & Reinstein, J. (2008). The Influence of Proline Isomerization and Off-Pathway Intermediates on the Folding Mechanism of Eukaryotic UMP/CMP Kinase. *J Mol Biol* **381**, 443-455.

55. Greene, L. H., Wijesinha-Bettoni, R. & Redfield, C. (2006). Characterization of the Molten Globule of Human Serum Retinol-Binding Protein Using NMR Spectroscopy. *Biochemistry* **45**, 9475-9484.

4

Non-native, long-range, hydrophobic interactions exist in unfolded apoflavodoxin

Sanne M. Nabuurs, Bregje de Kort, Adrie H. Westphal and Carlo P. M. van Mierlo

Abstract

Transient structures in unfolded proteins are key in the elucidation of the molecular details of initiation of protein folding. Recently, non-native secondary structure has been revealed in the unfolded state of *A. vinelandii* flavodoxin. These structured elements transiently interact and subsequently form the ordered core of an off-pathway folding intermediate. To investigate long-range interactions in unfolded apoflavodoxin that lead to formation of this off-pathway intermediate, use is made in this study of site-directed spin labeling. For this purpose, glutamine at position 48, which resides in a non-native α-helix of unfolded apoflavodoxin, is replaced by cysteine. This replacement enables covalent attachment of different nitroxide spin labels, such as MTSL and CMTSL. Due to this amino acid replacement stability of native apoflavodoxin against unfolding decreases and attachment of the nitroxide spin label MTSL leads to a further decrease in stability. Replacement of Gln48 by Cys48 decreases flexibility of the ordered regions in unfolded apoflavodoxin in 3.4 M GuHCl, due to increased hydrophobic interactions. Non-specific hydrophobic interactions between nitroxide spin labels and hydrophobic patches of unfolded apoflavodoxin perturb the unfolded protein and complicate interpretation of the spin label data obtained. Thus, care needs to be taken in the use of spin labels for the study of the conformational and dynamic properties of unfolded proteins. The spin label data show that non-native contacts exist between transiently ordered structured elements in unfolded apoflavodoxin.

Introduction

The understanding of the molecular mechanisms of proteins folding is one of the fundamental challenges of structural biology. Proteins initially fold form disordered unfolded states. Transient structures in unfolded proteins are key in the elucidation of the molecular details of initiation of protein folding. Early observations of residual structure in unfolded proteins prompted the suggestion that native-like structure is present in the denatured state [1-3] and that this residual structure biases the subsequent conformational search toward the native conformation [3,4]. However, evidence of non-native structure is found in unfolded states of a few proteins [5-8].

Recently, non-native secondary structure and non-native hydrophobic interactions have been observed in the unfolded state of a 179-residue flavodoxin from *Azotobacter vinelandii* [9], the protein of interest in the study presented here. Heteronuclear NMR data showed that structure formation in unfolded flavodoxin doesn't direct folding to the native state, but instead causes the formation of a misfolded off-pathway intermediate [10].

To investigate long-distance interactions in unfolded proteins paramagnetic relaxation agents are often used [11]. Site-specific labeling with thiol-specific nitroxide electron spin-labels, such as MTSL ((1-oxy-,2,2,5,5-tetramethyl-D-pyrroline-3-methyl)-methanethiosulfonate) [12] or CMTSL ((1-oxy-,2,2,5,5-tetramethylpyrroline-3-yl)-carbamidoethyl methanethiosulfonate) [13] (Fig. 4.1), allows determination of long-range contacts in unfolded proteins [14]. MTSL can exist in two electronic states, an oxidized, paramagnetic state and a reduced, diamagnetic state. In the oxidized state an unpaired electron in MTSL reduces the intensity of NMR signals of resonances that are within approximately 25 Å of the spin-label. This phenomenon is called paramagnetic relaxation enhancement (PRE).
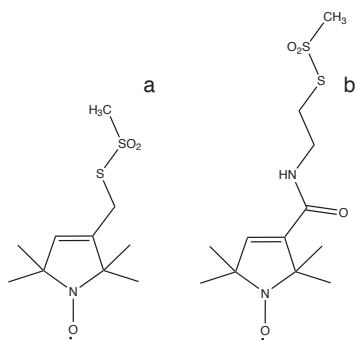
Figure 4.1 Schematic representations of the chemical structures of (a) MTSL and (b) CMTSL. Compared to MTSL, CMTSL contains an additional amide group in the linker between protein and the PROXYL ring, thereby introducing a degree of polarity into an otherwise hydrophobic compound.
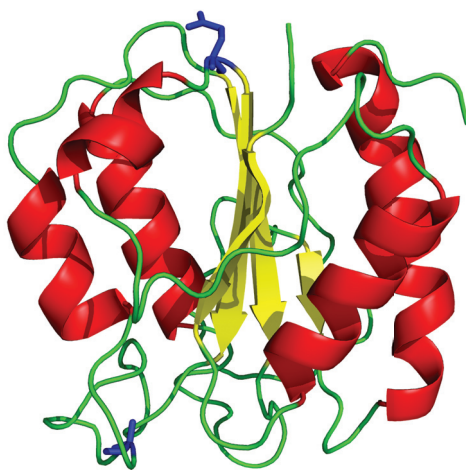
Figure 4.2 Cartoon drawing of native C69A flavodoxin from *A. vinelandii* (pdb ID 1YOB [39]). The protein adopts the flavodoxin-like or α-β parallel topology and is characterized by a parallel β-sheet surrounded by α-helices at either side of the sheet. The FMN cofactor is not shown. Residues Ala69 and Gln48 are shown in blue in stick representation.

Regions of a polypeptide chain that form tertiary interactions with the spin-labeled region will exhibit strong PRE effects. Conversely, regions that remain distant at all times should exhibit weak PRE effects.

Flavodoxins are monomeric proteins involved in electron transport, and contain a non-covalently bound flavin mononucleotide (FMN) cofactor. The proteins consist of a single structural domain and adopt the flavodoxin-like or α-β parallel topology, which is widely prevalent in nature (Fig. 4.2). Both denaturant-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin (i.e., flavodoxin without FMN) have been characterized using guanidine hydrochloride (GuHCl) as denaturant [15-20]. The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to flavodoxin except that residues in the flavin-binding region of the apo protein have considerable dynamics [21, 22]. In presence of FMN, binding of FMN to native apoflavodoxin is the last step in flavodoxin folding.

Apoflavodoxin kinetic folding involves an energy landscape with two folding intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [17]. Intermediate I_{on} lies on the productive route from unfolded to native protein, is highly unstable and is therefore not observed during denaturant-induced equilibrium unfolding. Consequently, GuHCl-induced equilibrium unfolding of apoflavodoxin is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin [17]. Approximately 90% of folding molecules fold via molten globule-like off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [17]. The formation of an off-pathway species is typical for proteins with a flavodoxin-like topology [19]. An off-pathway intermediate is experimentally observed for all other α-β parallel proteins of which the kinetic folding has been investigated, i.e., apoflavodoxin from *Anabaena* [23], CheY [24], cutinase [25, 26] and UMP/CMP kinase [27].

To better understand why the off-pathway intermediate is formed during flavodoxin folding, GuHCl-unfolded apoflavodoxin has been characterized at the residue-level using heteronuclear NMR spectroscopy [9,10]. Secondary shifts analysis and investigation of ^1H-^{15}N relaxation rates revealed four structured elements that transiently exist in unfolded apoflavodoxin. These transiently ordered regions have restricted flexibility on the (sub)nanosecond timescale and comprise residues Ala41 – Gly53, Glu72 – Gly83, Gln99 – Ala122 and Thr160 – Gly176 [9]. These regions match with regions of large average area buried upon folding (AABUF), which correlates with hydrophobicity [30], and corresponds to sequence-dependent dynamic variations due to hydrophobic interactions in unfolded proteins [28, 29]. This restricted flexibility is due to transient helix formation and local and non-local hydrophobic interactions [9, 10]. Upon decreasing denaturant concentration, the four structured elements in unfolded apoflavodoxin transiently interact and subsequently form the ordered core of the molten globule [9, 10]. Structure formation within virtually all parts of unfolded apoflavodoxin precedes folding to the molten globule state. This folding transition is non-cooperative and involves a series of distinct transitions [10]. Part of I_{off} remains random coil down to a GuHCl concentration of 1.58 M (i.e., residues Lys13 to Val36).

To probe long-range interaction in unfolded apoflavodoxin and to consolidate the findings of non-native interactions that lead to formation of the off-pathway intermediate, in this study use is made of site-directed spin labeling. To enable attachment of a maleimide spin label, Glu48, which resides in the non-native α-helix that is formed in unfolded apoflavodoxin [9], is replaced by cysteine, resulting in the Q48C apoflavodoxin variant. Note that wild-type (WT) flavodoxin contains a single cysteine at position 69. This cysteine is not accessible to solvent in the holo form of the protein, due to presence of FMN. However, in apoflavodoxin it is solvent accessible and can be used to attach

a spin label. In native apoflavodoxin Gln48 and Cys69 are positioned at two different sites of the protein (Fig. 4.2). Here, we show that attachment of a hydrophobic spin label leads to non-specific hydrophobic interactions between spin label and residues of the unfolded protein. Despite these interactions, still valid information about residual structure in unfolded apoflavodoxin is obtained.

Materials and Methods

Sample preparation. *A. vinelandii* (strain ATCC 478) flavodoxin II contains a single cysteine at position 69. Site-directed mutagenesis was used to replace the cysteine in the wild-type protein (WT) by alanine (C69A) and to replace glutamine at position 48 by cysteine resulting in the double mutant C69A-Q48C flavodoxin, hereafter named Q48C flavodoxin. Uniformly ^{15}N labeled Q48C and C69A flavodoxin was obtained from transformed *E. coli* cells grown on ^{15}N labeled algae medium (Silantes, Germany). Uniformly ^{15}N-labeled WT flavodoxin was obtained from transformed *E. coli* cells grown on ^{15}N-labeled minimal medium. All protein variants were purified as described [15].

Unfolded apoflavodoxin was obtained by denaturing flavodoxin in 6 M GuHCl. Subsequently, FMN was removed via gel filtration at 5 M GuHCl. WT and Q48C variants were labeled with nitroxide spin labels MTSL (1-oxy-,2,2,5,5-tetramethyl-D-pyrroline-3-methyl)-methanethiosulfonate or CMTSL (1-oxy-,2,2,5,5-tetramethylpyrroline-3-yl) carbamidoethyl methanethiosulfonate (Toronto Research Chemicals, Toronto) by adding a three-fold molar excess of spin label to unfolded protein at room temperature. After two hours, unreacted spin label was separated from labeled protein via gel filtration. Spin labeled WT and Q48C apoflavodoxin are hereafter called WT$^{(C)MTSL}$ and Q48C$^{(C)MTSL}$, respectively.

All NMR samples contained about 0.3 - 0.5 mM apoflavodoxin, 10% D_2O and 2,2-Dimethyl-2-silapentane-5-sulfonic acid as internal chemical shift reference. A sufficient amount of DTT to avoid covalent dimerisation of protein molecules was present in unlabeled samples of WT and Q48C apoflavodoxin. Refractometry was used to verify GuHCl concentration [31]. The buffer used was 100 mM potassium pyrophosphate (KPPi), pH 6.0.

NMR experiments. Gradient enhanced ^{1}H-^{15}N HSQC spectra were recorded on a Bruker Avance 700 MHz machine. Sample temperature was 25 °C. In the ^{1}H dimension of the ^{1}H-^{15}N HSQC experiments, 2048 complex data points were acquired, whereas in the indirect ^{15}N dimension 360 complex data points were collected. Spectral widths were 6010 and 1750 Hz in t_2 and t_1, respectively and a total of 16 scans was performed. All NMR experiments of spin labeled protein were repeated with a three-fold molar excess of ascorbic acid, by adding 5 μL from a concentrated stock (dilution < 0.1%).

Fluorescence Spectroscopy. Thermal unfolding of C69A apoflavodoxin, Q48C apoflavodoxin and Q48C^{MTSL} apoflavodoxin was followed by fluorescence emission. Protein unfolding was achieved by increasing temperature in a 1.5 ml stirred quartz cuvette (path-length 0.4 cm) from 20 to 70 °C at a rate of 0.5 °C/min. Temperature was measured in the cuvette by using an internal probe. The protein was excited at 280 nm, and fluorescence emission was recorded at 340 nm. The excitation and emission slits were set at 5 nm. In thermal unfolding experiments, protein was in 100 mM KPPi, pH 6.0, and protein concentration ranged between 4 - 6 μM.

Data analysis.
(a) NMR data. All NMR spectra were processed with NMRPipe [32] and analyzed using NMRViewJ [33]. Paramagnetic relaxation enhancement by nitroxide spin labels in the unfolded apoflavodoxin variants were measured as ratios of maximal cross peak intensities between ^{1}H-^{15}N HSQC spectra of unfolded WT$^{(C)MTSL}$ or Q48C$^{(C)MTSL}$ apoflavodoxin in the paramagnetic and diamagnetic state:

$$\text{Intensity ratio} = I_{para}/I_{dia} \qquad [4.1]$$

(b) Thermal-induced Equilibrium Unfolding Data. The change in free energy for thermal-induced protein unfolding, $\Delta G(T)$, is described by the modified Gibbs-Helmholtz equation:

$$\Delta G(T) = \Delta H_m\left(1 - T/T_m\right) - \Delta C_p\left[\left(T_m - T\right) + T\ln\left(T/T_m\right)\right] \qquad [4.2]$$

where ΔH_m is the enthalpy change for unfolding measured at T_m, T is the absolute temperature, T_m, the temperature at the midpoint of the transition, and ΔC_p is the difference in heat capacity between the unfolded and folded states [34, 35]. Under the assumption that ΔC_p is temperature independent [36], a two-state mechanism of unfolding can be fitted to individual thermal unfolding curves:

$$Y_{obs} = \left(a_U + b_U T\right) + \frac{\left(a_N + b_N T\right) - \left(a_U + b_U T\right)}{1 + e^{\frac{-\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) + \frac{\Delta C_p}{R}\left(\left(\frac{T_m}{T} - 1\right) + \ln\frac{T}{T_m}\right)}} \qquad [4.3]$$

where Y_{obs} is the measured far-UV CD or fluorescence signal, R is the gas constant, and a and b the intercepts and slopes of the pre- and post unfolding baselines, respectively.

Results and Discussion

Introducing a cysteine at position 48 destabilizes native apoflavodoxin. Thermal-induced unfolding experiments show that replacement of Gln48 by Cys48 destabilizes native apoflavodoxin in 100 mM KPPi, pH 6.0. Using fluorescence emission, the thermal midpoint of unfolding (T_m) of Q48C apoflavodoxin is determined to be 40.4 ± 0.2 ºC, whereas C69A apoflavodoxin has a T_m of 48.2 ± 0.1 ºC (Fig. 4.3). Attachment of MTSL spin label to Cys-48 further decreases T_m to 33.0 ± 1.6 ºC. Fluorescence emission monitors unfolding of native protein molecules, as both folding intermediate and unfolded apoflavodoxin have comparable fluorescence signals.

Introduction of a cysteine at position 48 changes the dynamic features of unfolded apoflavodoxin. The comparison of chemical shifts of cross peaks in ^{1}H-^{15}N HSQC spectra of Q48C and C69A apoflavodoxin in 3.4 M GuHCl showed that replacement of a glutamine by a cysteine at position 48 leads to chemical shift changes in unfolded apoflavodoxin. These chemical shift changes indicate long-range non-native interactions between transiently formed helices in unfolded apoflavodoxin [9].

In 3.4 M GuHCl, compared to the cross peak intensities of the backbone amides of unfolded C69A apoflavodoxin, several cross peak intensities of Q48C apoflavodoxin are decreased. This cross peak broadening is caused by reduced flexibility of backbone amides. Figure 4.4 shows that

four distinct regions in unfolded Q48C apoflavodoxin have reduced flexibility of their backbone amides compared to the corresponding amides of C69A apoflavodoxin. These distinct regions coincide with the four regions with restricted flexibility detected in unfolded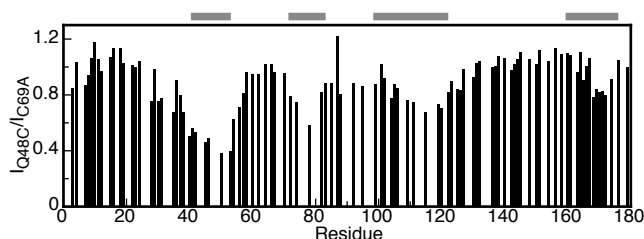 C69A apoflavodoxin [9], which are highlighted by gray bars in Figure 4.4. Consequently, upon replacing Gln48 by Cys48 the four regions with restricted flexibility in unfolded apoflavodoxin become less flexible. The latter is likely caused by non-native hydrophobic interactions between Cys48 and the four ordered regions identified.
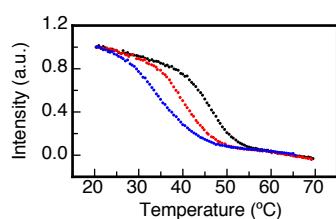
Figure 4.3 Thermal unfolding of apoflavodoxin shows that replacement of Gln48 by Cys48 and subsequent attachment of MTSL spin label to Cys48 both destabilize the protein against thermal unfolding. Thermal unfolding of C69A apoflavodoxin (black dots), Q48C apoflavodoxin (red dots) and Q48C[MTSL] apoflavodoxin (blue dots), as measured by changes in fluorescence emission at 340 nm. Midpoints of unfolding are 48.2 ± 0.1 °C for C69A apoflavodoxin, 40.4 ± 0.2 °C for Q48C apoflavodoxin and 33.0 ± 1.6 °C for Q48C[MTSL] apoflavodoxin.

Figure 4.4 Upon replacing Gln48 by Cys48 the four regions with restricted flexibility in unfolded apoflavodoxin in 3.4 M GuHCl become less flexible. Shown are the ratios between cross peak intensities of backbone amides of unfolded Q48C apoflavodoxin (I_{Q48C}) and unfolded C69A apoflavodoxin (I_{C69A}), which are both in 3.4 M GuHCl. Ratios are normalized to 1 for residue 179. Horizontal gray bars highlight the four regions of unfolded C69A apoflavodoxin with restricted flexibility on the (sub)nanosecond timescale [9].
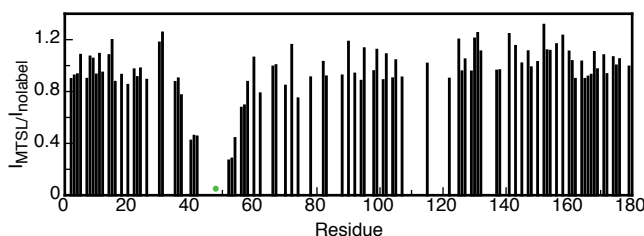
Figure 4.5 Attachment of MTSL to unfolded Q48C apoflavodoxin in 6.0 M GuHCl causes the protein to behave more ordered than random coil apoflavodoxin. Shown are the ratios between cross peak intensities of backbone amides of unfolded Q48C[MTSL] apoflavodoxin (I_{MTSL}) and unfolded Q48C apoflavodoxin ($I_{nolabel}$), which are both in 6.0 M GuHCl. Ratios are normalized to a value of 1 for residue 179. The green dot indicates the position of the MTSL spin label.

Attachment of MTSL to unfolded Q48C apoflavodoxin in 6.0 M GuHCl causes the protein to behave more ordered than random coil apoflavodoxin. Far-UV CD data and transverse relaxation rates [9] show that C69A apoflavodoxin in 6.0 M GuHCl behaves as a random coil. The Q48C variant of apoflavodoxin is also a random coil at 6.0 M GuHCl, as the ^{1}H-^{15}N HSQC spectra of both Q48C and C69A apoflavodoxin at this concentration denaturant are very similar regarding cross peak positions and intensities.

In Figure 4.5 the ratios between cross peak intensities of reduced Q48C^{MTSL} and Q48C apoflavodoxin, both unfolded in 6.0 M GuHCl, are shown. The direct sequential neighbours of Cys48 are either not visible, or could not be assigned, in the ^{1}H-^{15}N HSQC spectrum of Q48C^{MTSL} apoflavodoxin. Residues Ser40, Ala41, Glu42, Ile51, Leu52 and Gly53 have decreased intensities of their backbone amides in the HSQC spectrum of Q48C^{MTSL} apoflavodoxin in 6.0 M GuHCl compared to the intensities observed in the corresponding HSQC spectrum of Q48C apoflavodoxin (Fig. 4.5). This decrease in intensities is most likely caused by motional restrictions of these residues due to hydrophobic interactions with MTSL, which is hydrophobic of nature.

In conclusion, attachment of MTSL to unfolded Q48C apoflavodoxin in 6.0 M GuHCl causes the protein to behave more ordered than random coil apoflavodoxin would. In contrast, attachment of MTSL to Cys69 of WT apoflavodoxin in 6.0 M GuHCl does not affect the random coil behavior of the protein (data not shown).

Use of PRE experiments. Interaction between a nitroxide spin-label and a nearby proton causes broadening of the corresponding NMR signal because of an increased transverse relaxation rate of the proton involved [1]. This increase has an r^{-6} dependence on electron-proton distance and thus allows detection of long-range interactions in proteins. Consequently, ratios of cross peak intensities
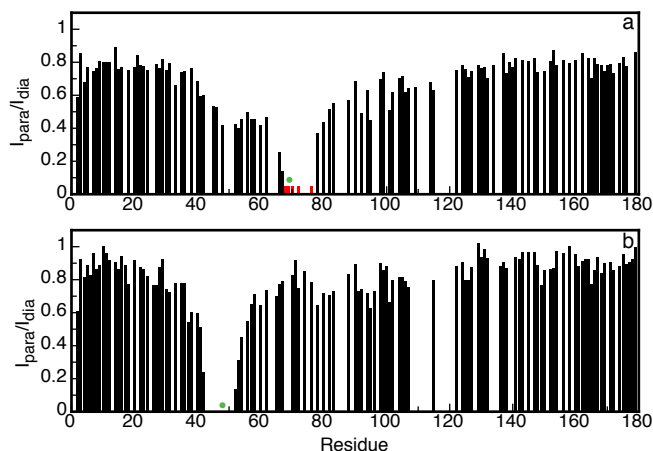

Figure 4.6 PRE of amide protons of WT^{MTSL}- and Q48C^{MTSL}-apoflavodoxin, which are unfolded in 6.0 M GuHCl. ^{1}H-^{15}N HSQC spectra of both protein variants with the spin label either in the paramagnetic or diamagnetic state have been recorded at 25 °C. Subsequently, the ratio of the intensities of cross peak maxima (I_{para}/I_{dia}) of backbone amides is determined. Shown are I_{para}/I_{dia} of (a) WT^{MTSL} and (b) Q48C^{MTSL} apoflavodoxin. Green dots highlight the positions of the MTSL spin label. Residues for which no cross peak is visible in the HSQC spectrum of the protein with the spin label in the paramagnetic state, but for which cross peaks are observed when the spin label is in the diamagnetic state are colored red in (a).

extracted from two redox state-dependent ^{1}H-^{15}N HSQC spectra of the spin-labeled protein, i.e., obtained in presence and in absence of the nitroxide radical (I_{para}/I_{dia}), permit estimation of distances between the spin label and affected protons in the protein [38]. Because of the ubiquitous backbone fluctuations in unfolded apoflavodoxin [9], we did not attempt to extract quantitative distances from the data. Rather, a trend in PRE as a function of primary structure can be observed, as shown in the following.

In 6.0 M GuHCl, WT^{MTSL} apoflavodoxin is less random coil than C69A apoflavodoxin, which is due to non-specific hydrophobic interactions. ^{1}H-^{15}N HSQC spectra are acquired of WT^{MTSL}- and Q48C^{MTSL}-apoflavodoxin unfolded in 6.0 M GuHCl with the spin label either in the paramagnetic or diamagnetic state. Subsequently, the ratio of the intensities of cross peak maxima (I_{para}/I_{dia}) of backbone amides is determined. Figure 4.6 shows these ratios and demonstrates that for both protein variants resonances of residues that are sequential neighbours of the cysteine with a nitroxide radical are broadened beyond detection. In addition, in case of WT^{MTSL} apoflavodoxin the nitroxide radical also decreases the backbone amide cross peak intensities of residues Ser40 - Leu62. PRE data of WT^{MTSL} apoflavodoxin thus indicate that residue 48 is in vicinity of the spin label that is attached to residue 69.

In contrast to the above observation, PRE data Q48C^{MTSL} apoflavodoxin (Fig. 4.6b) show no evidence of an interaction between Cys48 and Ala69. The nitroxide radical of MTSL attached to Cys48 does not broaden resonances of residues in the vicinity of Ala69 (unfortunately, the cross peak of the backbone amide of Ala69 suffers from severe overlap and is thus not assigned). The interaction between the MTSL spin label attached to Cys-69 and region Ser40 - Leu62 of unfolded apoflavodoxin in 6.0 M GuHCl is thus due to non-specific hydrophobic interactions. Non-specific broadening effects of an MTSL spin label were also observed in dimerization studies of ARNT PAS-B [13]. Our observations show that in 6.0 M GuHCl WT^{MTSL} apoflavodoxin is less random coil than C69A apoflavodoxin, as was observed for Q48C^{MTSL} apoflavodoxin in 6.0 M GuHCl.

In both WT^{CMTSL} and Q48C^{CMTSL} apoflavodoxin the spin label interacts with four transiently structured regions in unfolded protein in 3.4 M GuHCl. To reduce non-specific hydrophobic interactions in unfolded apoflavodoxin another spin label, i.e., CMTSL [13], which is more hydrophilic than MTSL, is chosen by us as paramagnetic relaxation agent. Compared to MTSL, CMTSL contains an additional amide group in the linker between protein and the PROXYL ring (see Fig. 4.1), thereby introducing a degree of polarity into an otherwise hydrophobic compound.

Figures 4.7a and 4.7b show the ratios of the intensities of cross peak maxima (I_{para}/I_{dia}) of backbone amides of the paramagnetic and diamagnetic states of WT^{CMTSL} apoflavodoxin and Q48C^{CMTSL} apoflavodoxin, both unfolded in 6.0 M GuHCl. The pattern of interactions of the CMTSL spin label in both unfolded proteins is similar to the interactions observed using MTSL as spin label. Consequently, also with CMTSL attached to unfolded apoflavodoxin in 6.0 M GuHCl, non-specific interactions exist between the spin label at position 69 and the hydrophobic region comprising residues 40 – 60 of unfolded apoflavodoxin.

As previously demonstrated, in unfolded apoflavodoxin in 3.4 M GuHCl transiently ordered structural elements exist [9]. Analysis of chemical shift differences between backbone amides of unfolded C69A and Q48C apoflavodoxin showed that non-native contacts exist between these transiently ordered structured elements [9]. PRE data of unfolded apoflavodoxin should strengthen these findings. In Figure 4.7c and 4.7d PRE data of WT^{CMTSL} and Q48C^{CMTSL} apoflavodoxin in 3.4 M

GuHCl are shown. All cross peaks of the backbone amides of WT[CMTSL] apoflavodoxin in the reduced state are also visible in the HSQC spectrum of WT[CMTSL] apoflavodoxin in the paramagnetic oxidised state. In both WT[CMTSL] and Q48C[CMTSL] apoflavodoxin the CMTSL spin label has interactions with various regions of the unfolded protein. In particular, interactions are observed between the spin label and the four transiently structured regions in unfolded apoflavodoxin in 3.4 M GuHCl, strengthening our previous findings.

In case of Q48C[CMTSL] apoflavodoxin unfolded in 3.4 M GuHCl, the spin label interacts with residues spread over the entire sequence of the protein. Residues 40 - 54 and Leu78 stand out, as the corresponding backbone amide resonances are broadened beyond detection. These residues will be discussed in the following. In addition, residues 82 to 129 are more severely broadened than the majority of residues of unfolded Q48C[CMTSL] apoflavodoxin.

The attached spin label induces the presence of two distinct states in unfolded apoflavodoxin.
We notice that in the HSQC spectrum of reduced Q48C[CMTSL] apoflavodoxin in 3.4 M GuHCl, two cross peaks are observed for each individual backbone amide of residues Ser40, Ala41, Glu42, Gln46, Phe49, Leu50 and Leu52 Gly53, Thr54 and Leu78. In Figure 4.8 examples of this cross peak doubling are shown. For each doubled cross peak it is observed that upon bringing the spin label into the paramagnetic state one of these two cross peaks broadens beyond detection. Doubling of the cross peaks of the residues mentioned is not due to presence of protein molecules with no spin label attached to them. In the latter case, additional cross peaks would appear at different positions in the spectrum, as the HSQC spectrum of Q48C apoflavodoxin in 3.4 M GuHCl shows (Fig. 4.8).

A plausible explanation for the observed doubling of cross peaks is that the attached spin label induces the presence of two distinct states in unfolded apoflavodoxin. In one of these states the spin label is in proximity of the above mentioned residues, whereas in the other state it is not. Both folding states are in slow exchange with one another on the NMR chemical shift time scale, as per backbone amide of the residues discussed two separate, sharp cross peaks are observed.

Leu78 is the only residue that is not sequentially close to the spin label at position 48 but does give rise to two backbone amide cross peaks in Q48C[CMTSL] apoflavodoxin in 3.4 M GuHCl. Indeed, chemical shift deviations upon replacing residue 48 showed that interactions between Cys48 and Leu78 must exist in unfolded apoflavodoxin [9]. In unfolded apoflavodoxin residual structure that is neither an α-helix nor a β-sheet was found in the region Glu72 – Gly83 [9]. These observations are further support for the existence of persistent hydrophobic interactions between Leu78 and the CMTSL spin label in one of the two distinct states within unfolded Q48C[CMTSL] apoflavodoxin. The tertiary interaction between residues 48 and 78 in unfolded apoflavodoxin, as revealed by PRE experiments, must be non-native, since in native flavodoxin the distance between the corresponding C^α atoms is 17.28 Å, as calculated using the X-ray structure of *A. vinelandii* flavodoxin [39].

Introducing cysteine residues in (unfolded) proteins to label them with hydrophobic probes has been used frequently [40, 41]. The results presented here show that in an unfolded protein, attachment of either a MTSL or CMTSL spin label to cysteine can introduce non-specific hydrophobic interactions.

MTSL is a widely used nitroxide spin label and is often employed to detect long-range interactions within (natively) unfolded proteins, such as apomyoglobin [2], α-synuclein [42], ACBP [38] and N-PGK [43]. As shown here for native apoflavodoxin, it has also been reported that the stability of ACBP variants decreases upon covalently attaching MTSL [38]. Remarkably, in contrast to unfolded apoflavodoxin, none of the mentioned unfolded proteins shows non-specific hydrophobic
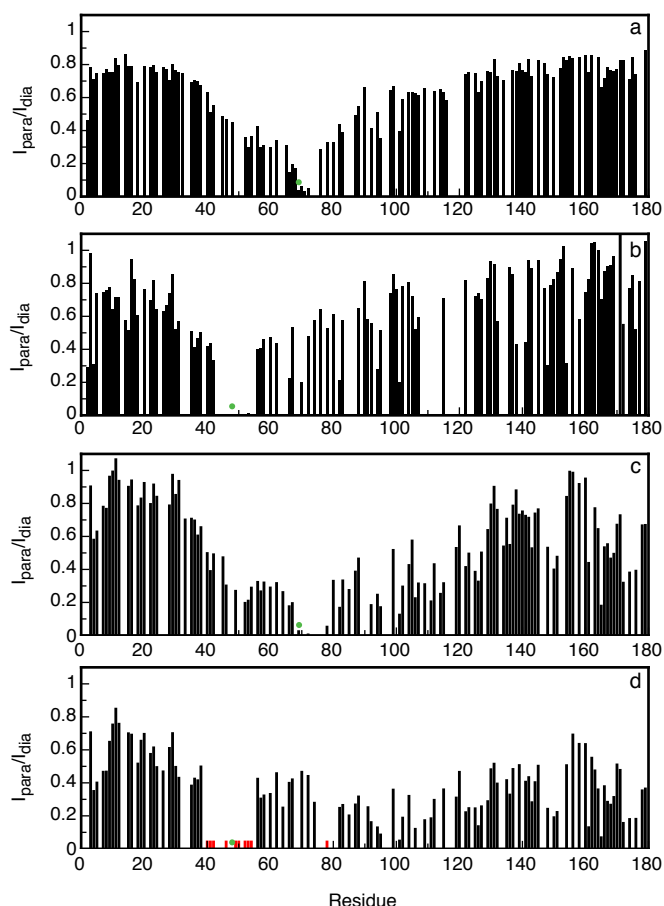
Figure 4.7 PRE of amide protons of WT[CMTSL]- and Q48C[CMTSL]-apoflavodoxin, which are unfolded in 6.0 M GuHCl. ^{1}H-^{15}N HSQC spectra of both protein variants with the spin label either in the paramagnetic or diamagnetic state have been recorded at 25 ºC. Subsequently, the ratio of the intensities of cross peak maxima (I_{para}/I_{dia}) of backbone amides is determined. Shown are I_{para}/I_{dia} of (a) WT[CMTSL] and (b) Q48C[CMTSL] apoflavodoxin, both in 6.0 M GuHCl, and I_{para}/I_{dia} of (c) WT[CMTSL] and (d) Q48C[CMTSL], both in 3.4 M GuHCl. Green dots highlight the positions of the CMTSL spin label . Residues Ser40, Ala41, Glu42, Gln46, Phe-49, Leu50, Leu52, Gly53, Thr54 and Leu78 of Q48C[CMTSL] apoflavodoxin in 3.4 M GuHCl, each give rise to two backbone amide cross peaks (as shown in Figure 4.8) and are indicated with red bars in (d).

interactions with the MTSL label. In addition, neither has doubling of cross peaks of these spin labeled unfolded proteins been observed. Possibly, the presence of many highly hydrophobic residues that form the parallel β-sheet of native apoflavodoxin, which is shielded from solvent, makes unfolded apoflavodoxin susceptible for non-specific interactions with nitroxide spin labels.
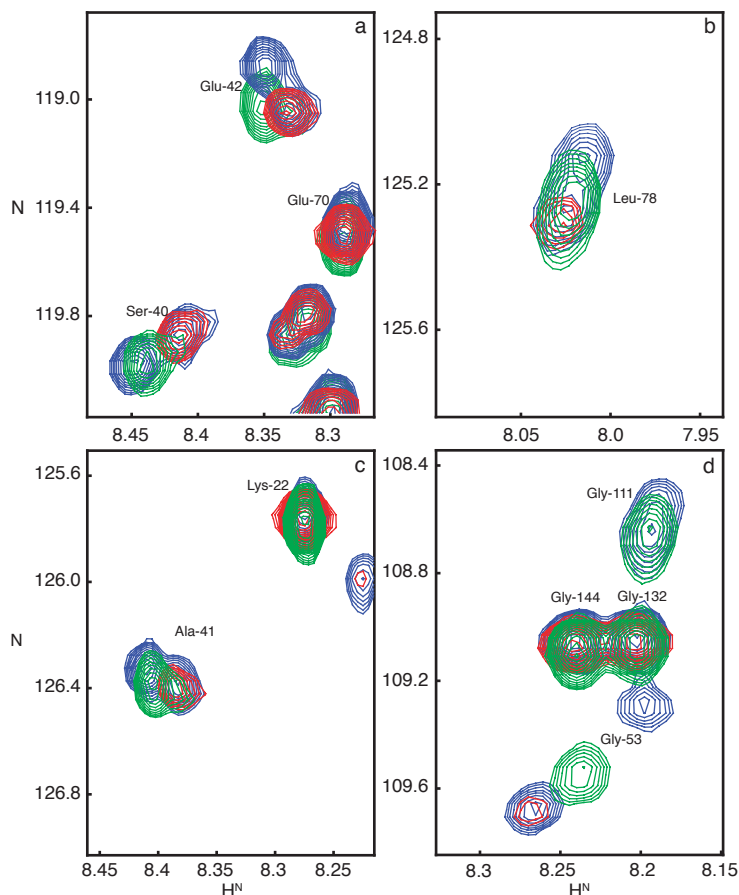
Figure 4.8 Backbone amides that are in close vicinity to the CMTSL spin label of Q48C^{CMTSL} apoflavodoxin in 3.4 M GuHCl can give rise to a set of two cross peaks in the corresponding ^{1}H-^{15}N HSQC spectrum. Blue contour levels highlight cross peaks arising from diamagnetic Q48C^{CMTSL} apoflavodoxin, red contour levels highlight cross peaks arising from paramagnetic Q48C^{CMTSL} apoflavodoxin, and green contour levels highlight cross peaks arising from Q48C apoflavodoxin with no spin label attached to it. A set two cross peaks is observed for (a) Ser40 and Glu42, (b) Leu78, (c) Ala41 and for (d) Gly53.

References

1. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* **268**, 158-169.

2. Lietzow, M. A., Jamin, M., Jane Dyson, H. J. & Wright, P. E. (2002). Mapping long-range contacts in a highly unfolded protein. *J Mol Biol* **322**, 655-662.

3. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J Mol Biol* **299**, 1341-1351.

4. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* **4**, 497-502.

5. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M. & Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J Mol Biol* **347**, 1053-1062.

6. Marsh, J. A., Neale, C., Jack, F. E., Choy, W. Y., Lee, A. Y., Crowhurst, K. A. & Forman-Kay, J. D. (2007). Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J Mol Biol* **367**, 1494-1510.

7. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2005). Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J Mol Biol* **346**, 279-294.

8. Reed, M. A., Jelinska, C., Syson, K., Cliff, M. J., Splevins, A., Alizadeh, T., Hounslow, A. M., Staniforth, R. A., Clarke, A. R., Craven, C. J. & Waltho, J. P. (2006). The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J Mol Biol* **357**, 365-372.

9. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. (2008). Extensive formation of off-pathway species during folding of an alpha-beta parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J Am Chem Soc* **130**, 16914-16920.

10. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. (2008). Non-cooperative formation of the off-pathway molten globule during folding of the α-β parallel protein apoflavodoxin. *J Am Chem Soc* **131**, 2739-2746.

11. Dyson, H. J. & Wright, P. E. (2005). Elucidation of the protein folding landscape by NMR. *Methods Enzymol* **394**, 299-321.

12. Berliner, L. J., Grunwald, J., Hankovszky, H. O. & Hideg, K. (1982). A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal Biochem* **119**, 450-495.

13. Card, P. B., Erbel, P. J. & Gardner, K. H. (2005). Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. *J Mol Biol* **353**, 664-677.

14. Mittag, T. & Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3-14.

15. van Mierlo, C. P. M., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Sci* **7**, 2331-2344.

16. van Mierlo, C. P. & Steensma, E. (2000). Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *J Biotechnol* **79**, 281-298.

17. Bollen, Y. J., Sanchez, I. E. & van Mierlo, C. P. M. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475-10489.

18. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem* **280**, 7836-7844.

19. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

20. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* **103**, 4095-4100.

21. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of *Azotobacter vinelandii* holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

22. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

23. Fernandez-Recio, J., Genzor, C. G. & Sancho, J. (2001). Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234-15245.

24. Kathuria, S. V., Day, I. J., Wallace, L. A. & Matthews, C. R. (2008). Kinetic traps in the folding of beta alpha-repeat proteins: CheY initially misfolds before accessing the native conformation. *J Mol Biol* **382**, 467-484.

25. Melo, E. P., Chen, L., Cabral, J. M., Fojan, P., Petersen, S. B. & Otzen, D. E. (2003). Trehalose favors a cutinase compact intermediate off-folding pathway. *Biochemistry* **42**, 7611-7617.

26. Otzen, D. E., Giehm, L., Baptista, R. P., Kristensen, S. R., Melo, E. P. & Pedersen, S. (2007). Aggregation as the basis for complex behaviour of cutinase in different denaturants. *Biochim Biophys Acta* **1774**, 323-333.

27. Lorenz, T. & Reinstein, J. (2008). The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP Kinase. *J Mol Biol* **381**, 443-455.

28. Le Duff, C. S., Whittaker, S. B. M., Radford, S. E. & Moore, G. R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding. *Journal of Molecular Biology* **364**, 824-835.

29. Schwarzinger, S., Wright, P. E. & Dyson, H. J. (2002). Molecular hinges in protein folding: The urea-denatured state of apomyoglobin. *Biochemistry* **41**, 12681-12686.

30. Rose, G. D. & Roy, S. (1980). Hydrophobic Basis of Packing in Globular-Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4643-4647.

31. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. In *Methods in enzymology* (Hirs, C. H. W. & Timasheff, S. N., eds.), Vol. 26, pp. 43-50. Academic Press, New York.

32. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* **6**, 277-293.

33. Johnson, B. A. & Blevins, R. A. (1994). NMR View - a Computer-Program for the Visualization and Analysis of NMR Data. *J Biomol NMR* **4**, 603-614.

34. Pace, C. N. & Laurents, D. V. (1989). A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-2525.

35. Becktel, W. J. & Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-1877.

36. Privalov, P. L. & Khechinashvili, N. N. (1974). A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J Mol Biol* **86**, 665-684.

37. van Mierlo, C. P. M., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

38. Teilum, K., Kragelund, B. B. & Poulsen, F. M. (2002). Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *J Mol Biol* **324**, 349-357.

39. Alagaratnam, S., van Pouderoyen, G., Pijning, T., Dijkstra, B. W., Cavazzini, D., Rossi, G. L., Van Dongen, W. M., van Mierlo, C. P. M., van Berkel, W. J. & Canters, G. W. (2005). A crystallographic study of Cys69Ala flavodoxin II from *Azotobacter vinelandii*: structural determinants of redox potential. *Protein Sci* **14**, 2284-2295.

40. Clore, G. M., Tang, C. & Iwahara, J. (2007). Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol* **17**, 603-616.

41. Fanucci, G. E. & Cafiso, D. S. (2006). Recent advances and applications of site-directed spin labeling. *Curr Opin Struct Biol* **16**, 644-653.

42. Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M. & Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Natl Acad Sci U S A* **102**, 1430-1435.

43. Cliff, M. J., Craven, C. J., Marston, J. P., Hounslow, A. M., Clarke, A. R. & Waltho, J. P. (2009). The denatured state of N-PGK is compact and predominantly disordered. *J Mol Biol* **385**, 266-277.

5

Topological switching between an α-β parallel protein and a remarkably helical molten globule

Sanne M. Nabuurs, Adrie H. Westphal, Marije aan den Toorn, Simon Lindhoud and Carlo P.M. van Mierlo (submitted)

Abstract

Partially folded protein species transiently exist during folding of most proteins. Often these species are molten globules, which may be on- or off-pathway to native protein. Molten globules have a substantial amount of secondary structure but lack virtually all tertiary side-chain packing characteristic of natively folded proteins. These ensembles of interconverting conformers are prone to aggregation and potentially play a role in numerous devastating pathologies, and thus attract considerable attention. The molten globule that is observed during folding of apoflavodoxin from *Azotobacter vinelandii* is off-pathway, as it has to unfold before native protein can be formed. Here we report that this species can be trapped under native-like conditions by substituting amino acid residue Phe44 for Tyr44, allowing characterization of its conformation. Whereas native apoflavodoxin contains a parallel β-sheet surrounded by α-helices (i.e., the flavodoxin-like or α-β parallel topology), it is shown that the molten globule has a totally different topology: it is helical and contains no β-sheet. The presence of this remarkably non-native species shows that single polypeptide sequences can code for distinct folds that swap upon changing conditions. Topological switching between unrelated protein structures is likely a general phenomenon in the protein structure universe.

Introduction

Four decades ago Levinthal predicted that folding of proteins to their native states involves folding pathways on which folding intermediates reside [1]. Today, the considerable advances made in experimental methods have shown that hardly any protein exhibits true two-state folding behavior [2]. Ultra-rapid mixing experiments led to the detection of short-lived kinetic intermediates [3] and techniques such as FRET, FCS and NMR revealed partially folded states that are difficult to detect [4, 5]. Occasionally proteins can be changed by mutagenesis to create a polypeptide that forms a relatively stable folding intermediate [6-8]. In general, experimental data of folding intermediates argue for near-native topology of these species, which have incompletely folded or partially misfolded structural elements [8-10].

Most folding proteins encounter folding energy landscapes [11] that are rough. As a result, partially folded intermediates, which may be on- or off-pathway to the native state, are populated. When the intermediate is on-pathway, as is observed for the majority of proteins studied to date, it has native-like topology and is productive for folding. In contrast, when the intermediate is off-pathway it is trapped in such a manner that the native state cannot be reached without substantial reorganizational events [2]. Several kinetic studies have revealed involvement of off-pathway intermediates during protein folding [12-18].

Structural characterization of folding intermediates is challenging due to the transient nature of their existence. In addition, if an intermediate is observed at equilibrium it is in most cases sparsely populated relative to the corresponding native and unfolded states. However, characterization of folding intermediates is crucial for the understanding of protein folding. Kinetic intermediates that appear early during folding have been shown to resemble the relatively stable molten globule [19-22] intermediates formed by several proteins under mildly denaturing conditions [19]. This resemblance has been demonstrated for α-lactalbumin [23], apomyoglobin [24], RNAse H [25], T4 lysozyme [26], and suggests that these molten globules can be considered as models of transient intermediates [2]. Due to exposed hydrophobic groups, molten globules are prone to aggregation. Particularly, a decrease of the folding rate due to the presence of an off-pathway molten globule, which is kinetically trapped and partially folded, increases the likelihood of protein aggregation. This aggregation phenomenon can have detrimental effects on organisms [28]. Understanding the formation and conformation of molten globules offers potential insights into factors responsible for protein misfolding, aggregation and, potentially, for numerous devastating pathologies [29].

Here, we report the trapping under native-like conditions of the molten globule folding intermediate of a 179-residue flavodoxin from *A. vinelandii*, enabling characterization of its conformation. Flavodoxins are monomeric proteins involved in electron transport, and contain a non-covalently bound FMN cofactor. These proteins consist of a single structural domain and adopt the flavodoxin-like or α-β parallel topology, which is widely prevalent in nature. Both denaturant-induced equilibrium and kinetic (un)folding of flavodoxin and apoflavodoxin (i.e., flavodoxin without FMN) have been characterized using guanidine hydrochloride (GuHCl) as denaturant [13, 30-34]. The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to flavodoxin with exception of residues in the flavin-binding region of the apo protein. These residues have considerable dynamics in apoflavodoxin [35, 36]. The last step in flavodoxin folding is binding of the cofactor to native apoflavodoxin [30].

Kinetic folding of apoflavodoxin involves an energy landscape with two intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [13]. Intermediate I_{on} lies on

the productive route from unfolded to native protein, is highly unstable and is therefore not observed during denaturant-induced equilibrium unfolding. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [13]. The formation of an off-pathway species is typical for proteins with a flavodoxin-like topology [32]. An off-pathway intermediate is experimentally observed for all other α-β parallel proteins of which the kinetic folding has been investigated: apoflavodoxin from *Anabaena* [12], CheY [16], cutinase [17], and UMP/CMP kinase [18].

Equilibrium unfolding of apoflavodoxin is described by the three-state model: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin (Fig. 5.1) [13]. Unfolded apoflavodoxin contains several transiently structured regions that dock, causing formation of I_{off} [37]. The intermediate folds in a non-cooperative manner [38] and populates significantly in the concentration range of 1 to 3 M GuHCl. Its maximal population is at 1.76 M GuHCl with a mole fraction of 0.63 (Fig. 5.1d). The off-pathway species is molten globule-like: it is compact, its three tryptophans are solvent exposed and it has severely broadened NMR resonances due to exchange between different conformers on the micro- to millisecond timescale [13, 33, 39]. Although the conformation of this molten globule is currently unknown and cannot be determined by using NMR spectroscopy, intermediate I_{off} probably contains helices [13]. Elevated protein concentrations [39] and molecular crowding [40] cause severe aggregation of this species.

Anfinsen's thermodynamic hypothesis states that each amino acid sequence codes for a protein which folds to the state that has the lowest free energy. This state is called the native state and is characterized by a unique tertiary fold [41]. In this study it is shown that the equilibrium between native and molten globule apoflavodoxin can be shifted, in absence of denaturant, from one monomeric species to the other. Full population of the molten globule folding state of apoflavodoxin is possible through covalent introduction of just a single extra oxygen atom in the protein, achieved by substituting Phe44 for Tyr44. This substitution leads to significant destabilization of native
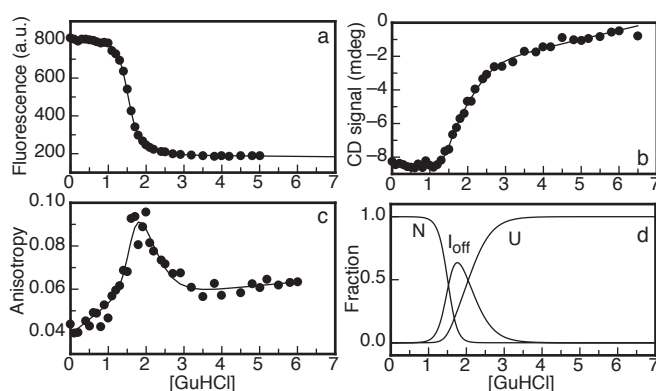

Figure 5.1 GuHCl-induced equilibrium (un)folding of *A. vinelandii* WT-apoflavodoxin shows the involvement of an intermediate during apoflavodoxin folding, as demonstrated by Bollen *et al.* [13]. (a) Fluorescence emission at 340 nm with excitation at 280 nm. (b) CD at 222 nm. (c) Fluorescence anisotropy data detected with a 335 nm cut-off filter, excitation is at 300 nm. The solid lines in panels a to c are the result of a global fit of a three-state model for equilibrium (un)folding [13]. (d) Normalized population of native (N), off-pathway intermediate (I_{off}), and unfolded (U) apoflavodoxin molecules as a function of denaturant concentration.

Tyr44-apoflavodoxin compared to WT-apoflavodoxin. Upon a mild change in conditions, i.e., lowering salt concentration, virtually all protein molecules exist as molten globule, as it has become the lowest energy species. Now, characterization of an off-pathway folding intermediate is possible in absence of denaturant. We show that it has a drastically different topology compared to native protein: it is helical and lacks the parallel β-sheet of native apoflavodoxin. Our observations show that interconversion between two unrelated protein-folds can be dictated by a single amino acid sequence.

Materials and Methods

Proteins. The single cysteine at position 69 in *A. vinelandii* (strain ATCC 478) flavodoxin II was substituted for an alanine to avoid covalent dimerization of apoflavodoxin. This protein variant is largely similar to wild-type flavodoxin [33, 42] and is referred to as WT-flavodoxin. Subsequently, in addition, the phenylalanine at position 44 was substituted for a tyrosine using site-directed mutagenesis. The latter protein variant is referred to as Tyr44-flavodoxin. Both flavodoxin variants were obtained from transformed *E. coli* cells, and purified as described [33].

To obtain native apoprotein, holoprotein was denatured in 6 M GuHCl. Subsequently, both FMN and denaturant were removed via gel filtration on a Superdex 75 column, which is loaded with 100 mM potassium pyrophosphate (KPPi), pH 6.0. During this removal step, the protein folds to native apoprotein.

Fluorescence spectroscopy. Steady-state fluorescence measurements of denaturant-induced equilibrium unfolding were done at 25 ºC on a Cary Eclipse fluorometer (Varian). Excitation was at 280 nm and emission was measured at 340 nm. Protein concentration was 4 μM in 100 mM KPPi at pH 6.0.

Fluorescence emission spectra were obtained at 25 ºC. Excitation was at 280 nm. Protein concentration was 4 μM. Buffer used was 100 mM KPPi, at pH 6.0, or a tenfold dilution of this buffer.

Thermal protein unfolding was followed by fluorescence emission. Temperature was increased in a 1.5 ml stirred quartz cuvette (path-length 0.4 cm) from 12 to 75 ºC at a rate of 1 ºC/min. Excitation was at 280 nm and emission was recorded at 340 nm. Protein was in 100 mM KPPi, pH 6.0, and protein concentration ranged between 2.5-3.0 μM.

In all experiments, excitation and emission slits were set to a width of 5 nm.

Far-UV CD spectroscopy. CD measurements were acquired using a Jasco J715 spectropolarimeter. Denaturant-induced equilibrium unfolding was followed at 25 ºC in a 2-mm quartz cuvette by measuring ellipticities at 222 and 255 nm, and the corresponding signal was averaged over 3 min/wavelength. The averaged ellipticity at 255 nm was subtracted from the averaged ellipticity at 222 nm. Protein concentration was 4 μM in 100 mM KPPi at pH 6.0.

CD spectra of protein in 1-mm quartz cuvettes were obtained by averaging 20 scans and corrected by subtracting spectra of corresponding blank solutions. Protein concentration ranged between 1 - 2 μM. Buffer used was 100 mM KPPi, at pH 6.0, or a tenfold dilution of this buffer.

Thermal unfolding of Tyr44-apoflavodoxin was followed at 210 nm. Temperature was increased from 10 to 75 °C at a rate of 1 °C/min in a 1.5 ml stirred quartz cuvette (path-length 0.4 cm). Protein concentration was 3 μM in 100 mM KPPi at pH 6.0.

Fluorescence Anisotropy. Anisotropy was measured at 25 °C using a Fluorolog 3.2.2 fluorometer (Horiba Jobin Yvon Ltd.). Excitation was at 300 nm with a 1 nm slit, emission was at 345 nm with a 14 nm slit. For each protein sample, 5 measurements were done at each setting and the data were subsequently averaged. In case of denaturant-induced equilibrium unfolding of Tyr44-apoflavodoxin, protein concentration was 4 μM in 100 mM KPPi, pH 6.0. In the experiments in which the KPPi concentration was decreased by diluting the buffer solution, protein concentration ranged between 2.7-3.0 μM.

NMR. Gradient-enhanced ^{1}H-^{15}N HSQC spectra were recorded on a Bruker AMX 500 MHz machine. Sample temperature was 25 °C.

Data analysis.
(a) Denaturant-induced Equilibrium Unfolding Data. Apoflavodoxin equilibrium denaturation data monitored by fluorescence emission intensity at 340 nm, far-UV CD at 222 nm, and fluorescence anisotropy at 345 nm were globally fitted to a three-state model for equilibrium denaturation (Equation 5.1), according to Equations 5.2-5.6, using ProFit (Quantum Soft, Zürich, Switzerland):

$$U \leftrightarrows I \leftrightarrows N \qquad\qquad\qquad\qquad [5.1]$$

$$K_{UI} = [I]/[U], \qquad K_{IN} = [N]/[I] \qquad\qquad [5.2]$$

$$K_{ij}(D) = K_{ij}{}^\circ \cdot exp(m_{ij}[D]) \qquad\qquad\qquad [5.3]$$

$$
\begin{aligned}
f_U &= 1/(1 + K_{UI} + K_{UI}K_{IN}) \\
f_I &= K_{UI}/(1 + K_{UI} + K_{UI}K_{IN}) \\
f_N &= K_{UI}K_{IN}/(1 + K_{UI} + K_{UI}K_{IN})
\end{aligned}
\qquad [5.4]
$$

$$Y^{obs} = (a_N + b_N[D])f_N + (a_I + b_I[D])f_I + (a_U + b_U[D])f_U \qquad [5.5]$$

In which U, I, and N represent the unfolded, intermediate, and native state of the protein, respectively, K_{ij} is the equilibrium constant of the *i-j* equilibrium, $K_{ij}{}^\circ$ is K_{ij} at zero denaturant concentration, [D] is the denaturant concentration, m_{ij} is the constant that describes the denaturant concentration-dependence of K_{ij}, f_i is the fractional population of state *i* at a particular denaturant concentration, Y^{obs} is the observed spectroscopic signal, a_i is the spectroscopic property of state *i* at zero concentration denaturant, and b_i is the constant describing the denaturant concentration-dependence of the spectroscopic signal of state *i*. Since not all Tyr44-apoflavodoxin molecules are native in 100 mM KPPi, pH 6.0 at 25 °C, constant a_N was fixed to 745 (a.u.), as this value corresponds to the fluorescence emission signal of 4 μM native WT-apoflavodoxin in the same buffer at 25 °C. The denaturant dependence of the fluorescence signal of native protein, b_N, was fixed to 0, as the fluorescence signal of native WT-apoflavodoxin hardly depends on denaturant concentration.

To incorporate fluorescence anisotropy data in the global analysis procedure, a modification of Equation 5.5 is required. Unlike fluorescence emission intensity and CD signals, fluorescence anisotropy signals do not linearly track the mole fractions of the three folding states of apoflavodoxin. Instead, fluorescence anisotropy signals are determined by the mole fraction of states involved, and by the fluorescence quantum yield of each individual state [43]. To describe the anisotropy signal adequately, the mole fraction of each species was multiplied with its intrinsic anisotropy (with corresponding denaturant concentration dependence) and weighted by its fluorescence intensity at 340 nm, according to:

$$Y_{anisotropy}^{obs} = \frac{(a_{N,Fl} + b_{N,Fl}[D])(a_{N,anis} + b_{N,anis}[D])f_N + (a_{I,Fl} + b_{I,Fl}[D])(a_{I,anis} + b_{I,anis}[D])f_I + (a_{U,Fl} + b_{U,Fl}[D])(a_{U,anis} + b_{U,anis}[D])f_U}{(a_{N,Fl} + b_{N,Fl}[D]) + (a_{I,Fl} + b_{I,Fl}[D]) + (a_{U,Fl} + b_{U,Fl}[D])}$$

[5.6]

with $a_{i,Fl}$ the intrinsic fluorescence intensity of species i at 340 nm, $b_{i,Fl}$ the denaturant concentration dependence of $a_{i,Fl}$, and $a_{i,anis}$ and $b_{i,anis}$ the intrinsic anisotropy of state i and its corresponding denaturant concentration dependence, respectively. As discussed, $a_{N,Fl}$, and $b_{N,Fl}$, were fixed to 745 and 0, respectively. Each individual data point was weighted by the square of the corresponding standard error during the global fit procedure.

(b) Thermal-induced Unfolding Data. The change in free energy for thermal-induced protein unfolding, $\Delta G(T)$, is described by the modified Gibbs-Helmholtz equation:

$$\Delta G(T) = \Delta H_m(1 - T/T_m) - \Delta C_p[(T_m - T) + T\ln(T/T_m)]$$ [5.7]

where ΔH_m, is the enthalpy change for unfolding measured at T_m, T is the absolute temperature, T_m, the temperature at the midpoint of the transition, and ΔC_p is the difference in heat capacity between the unfolded and folded states [44,45]. Under the assumption that ΔC_p is temperature independent [46], a two-state mechanism of unfolding can be fitted to individual thermal unfolding curves:

$$Y_{obs} = (a_U + b_U T) + \frac{(a_N + b_N T) - (a_U + b_U T)}{1 + e^{\frac{-\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right) + \frac{\Delta C_p}{R}\left(\left(\frac{T_m}{T} - 1\right) + \ln\frac{T}{T_m}\right)}}$$ [5.8]

where Y_{obs} is the measured far-UV CD or fluorescence signal, R is the gas constant, a and b the intercepts and slopes of the pre- and post unfolding baselines, respectively.

(c) Dissociation constant of the Tyr44-apoflavodoxin-FMN complex. The dissociation constant of the Tyr44-apoflavodoxin-FMN complex was determined using the quenching of FMN fluorescence upon binding of the cofactor to apoprotein. A solution of 1.5 ml containing 197 nM FMN in 100 mM potassium pyrophosphate, pH 6.0, was titrated with aliquots of 2.73 μM Tyr44-apoflavodoxin in the same buffer. After each addition of protein, the system was allowed to equilibrate for 5 min in the dark. Subsequently, FMN fluorescence intensity was determined. Excitation was at 445 nm with a slit of 5 nm, and emission was recorded at 525 nm with a slit of 10 nm. The dissociation constant of the Tyr44-apoflavodoxin-FMN complex was determined by fitting the fluorescence emission to Equation 5.9 [30],

$$F = dF_{end} + F_{\delta}\left(dC_F - \frac{\sqrt{(C_A + K_D + dC_F)^2 - 4C_A dC_F}}{2} \right) \qquad [5.9]$$

where F is the observed fluorescence intensity after each addition, d the dilution factor (initial volume/total volume), F_{end} the remaining fluorescence intensity after titration (resulting both from fluorescence of Tyr44-flavodoxin and from traces of flavin impurities that are unable to bind to Tyr44-apoflavodoxin), F_{δ} the difference in molar emission intensity between Tyr44-flavodoxin and free FMN, C_F the initial concentration of FMN, C_A the total protein concentration (i.e., apo + holo), and K_D the dissociation constant of the Tyr44-apoflavodoxin-FMN complex.

Results and Discussion

Design of Tyr44-apoflavodoxin. Residue Phe44 is a highly conserved amino acid residue in flavodoxins [47]. It is part of the major hydrophobic core of flavodoxin, containing residues that are in a rigid three-dimensional geometry in native protein (Fig. 5.2) [48]. Consequently, substitution of Phe44 for the slightly larger and more hydrophilic residue tyrosine is expected to decrease the stability of the native state of apoflavodoxin against unfolding. It potentially avoids significant population of the native state of Tyr44-apoflavodoxin under conditions that favor folding. This subtle residue substitution will have much less effect on the stability of the molten globule intermediate of Tyr44-apoflavodoxin than it has on the stability of native Tyr44-apoflavodoxin, as this intermediate with exposed hydrophobic groups is highly dynamic [13, 33, 39]. Unfolded apoflavodoxin is expected to remain largely unaltered upon substituting Phe44 for Tyr44, as unfolded molecules are more dynamic than molten globule protein molecules [37].

FMN is tightly bound to Tyr44-flavodoxin. The dissociation constant of the Tyr44-apoflavodoxin-FMN complex is determined by titrating a solution containing Tyr44-apoflavodoxin to a solution containing FMN (see Materials and Methods). Upon cofactor binding to apoflavodoxin FMN fluorescence quenches. The corresponding FMN binding curve is shown in Figure 5.3, and the fitted dissociation constant K_D of the Tyr44-apoflavodoxin-FMN complex is $(4.6 \pm 0.1) \cdot 10^{-10}$ M. This value
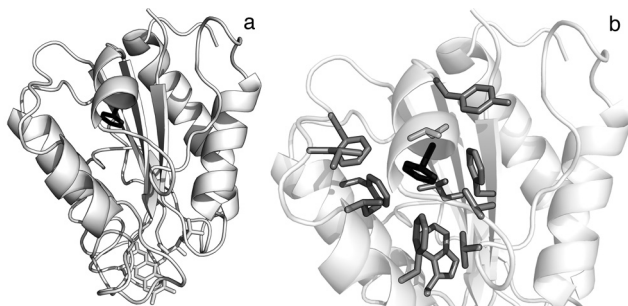

Figure 5.2 Residue Phe44 resides in a hydrophobic pocket of WT-flavodoxin. (a) Cartoon representation of WT-flavodoxin in which Phe44 (black) and FMN are shown in stick representation (pdb ID 1YOB). (b) Residues within 5 Å distance of Phe44 (black) are highlighted.
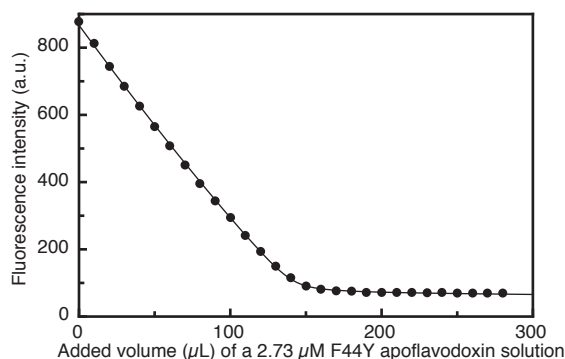
Figure 5.3 Determination of the dissociation constant of the Tyr44-apoflavodoxin-FMN complex using the quenching of FMN fluorescence upon FMN binding to Tyr44-apoflavodoxin. A 197 nM FMN solution was titrated with aliquots of a 2.73 μM Tyr44-apoflavodoxin solution. Equation 5.9 is fitted to the resulting fluorescence intensity data (see Materials and Methods). The dissociation constant is determined to be $(4.6 \pm 0.1) \cdot 10^{-10}$ M in 100 mM KPPi, pH 6.0, at 25 ºC.
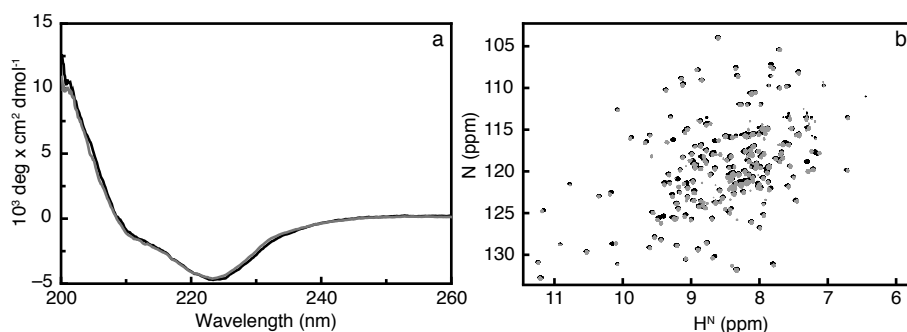

Figure 5.4 Tyr44-flavodoxin and WT-flavodoxin have nearly indistinguishable three-dimensional structures. (a) Far-UV CD spectra of WT-flavodoxin (black line) and of Y44-flavodoxin (gray line) demonstrate that the secondary structure content of both proteins is identical. (b) ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled WT-flavodoxin (black cross peaks) and Y44-flavodoxin (gray cross peaks) show that the microenvironments of the backbone amides of sequentially identical residues of both proteins are almost indistinguishable. Sample conditions are 100 mM KPPi, pH 6.0, at 25 ºC.

differs only slightly from the one that characterizes the WT-apoflavodoxin-FMN complex (i.e., K_D is $(3.4 \pm 0.6) \cdot 10^{-10}$ M [30]). Consequently, substitution of Phe44 for Tyr44 hardly affects the capacity of apoflavodoxin to bind FMN tightly.

Substitution of Phe44 for Tyr44 does not alter the conformation of (apo)flavodoxin. Tight FMN binding occurs primarily through a very specific combination and geometry of hydrogen bonds and aromatic interactions with apoflavodoxin [49]. Consequently, the observation that Tyr44-apoflavodoxin binds FMN almost as firmly as WT-apoflavodoxin, implies that the three-dimensional structures of Tyr44- and WT-flavodoxin must be similar. Indeed, secondary structure content of both proteins is identical, as revealed by far-UV CD spectroscopy (Fig. 5.4a). In addition, the ^{1}H-^{15}N HSQC spectra of both holoproteins are alike (Fig. 5.4b). Only a few, relatively small,

chemical shift changes due to changing the chemical identity of the side chain of residue 44 are observed. As HSQC spectra are fingerprints of protein conformations, the similarity in cross peak positions of the backbone amides of Tyr44-flavodoxin and WT-flavodoxin implies that the corresponding three-dimensional structures are nearly indistinguishable.

Just as observed for flavodoxin, it is expected that introduction of only a single extra oxygen atom into the protein, achieved by substituting Phe44 for Tyr44, does not affect the three-dimensional structure of native apoflavodoxin. This structure contains three tryptophans that have specific relative positions and orientations in the hydrophobic core of the molecule [48]. Indeed, in 100 mM KPPi, pH 6.0, at 25 °C, both protein variants give rise to almost identical fluorescence emission spectra (wavelengths of emission maxima are identical, i.e., 330 nm, and corresponding fluorescence emission intensities differ only slightly). Most importantly, under this experimental circumstance, native WT-apoflavodoxin has a strikingly low fluorescence anisotropy of 0.035 ± 0.003 [13] and anisotropy of Tyr44-apoflavodoxin is comparably low (i.e., anisotropy is 0.038 ± 0.002). The low anisotropy of native WT-apoflavodoxin is due to rapid unidirectional FRET from W167 to W128, as both tryptophans are only 7 Å apart [34]. Consequently, the conformations of native WT- and native Tyr44-apoflavodoxin must be nearly identical.

Although the substitution of Phe44 for Tyr44 does not alter the conformation of native apoflavodoxin, it is expected that this substitution decreases the stability of native protein against unfolding, as is shown below.

Native Tyr44-apoflavodoxin has a relatively low stability against unfolding and unfolds according to a three-state model. Far-UV CD and fluorescence spectroscopy are used to follow GuHCl-induced equilibrium unfolding of Tyr44-apoflavodoxin (Fig. 5.5). In contrast to the apparent two-state unfolding behavior of WT-apoflavodoxin detected by fluorescence emission at 340 nm (Fig. 5.1a) [13], the corresponding unfolding curve of Tyr44-apoflavodoxin is biphasic (Fig. 5.5a). Both
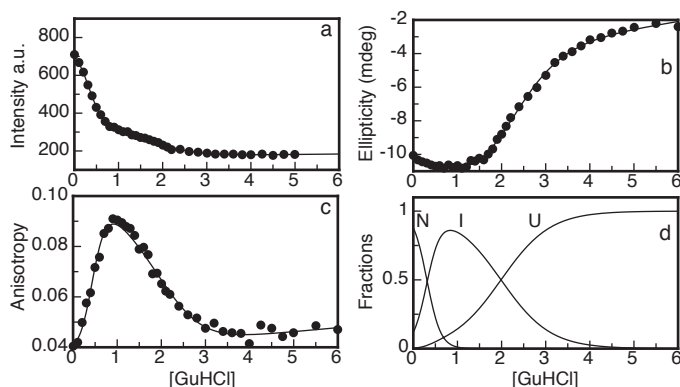

Figure 5.5 GuHCl-induced equilibrium (un)folding of Tyr44-apoflavodoxin shows that an intermediate folding state can be almost fully populated. (a) Fluorescence emission at 340 nm with excitation at 280 nm. (b) CD at 222 nm. (c) Fluorescence anisotropy data detected at 345 nm using a slit of 14 nm, excitation is at 300 nm. The solid lines in panels a to c are the result of a global fit of a three-state model for equilibrium (un)folding (see Materials and Methods). (d) Normalized population of native (N), folding intermediate (I), and unfolded (U) apoflavodoxin molecules as a function of denaturant concentration. Sample conditions are 100 mM KPPi, at pH 6.0.

far-UV CD and fluorescence anisotropy unfolding curves of Tyr44-apoflavodoxin resemble the corresponding unfolding curves of WT-apoflavodoxin. Population of a folding intermediate during Tyr44-apoflavodoxin folding is further substantiated by the observation that the unfolding curves detected by fluorescence emission and CD do not coincide (Fig. 5.5). In addition, the unfolding curve obtained by fluorescence anisotropy is biphasic.

The global fit of a three-state folding model to the three unfolding data sets presented in Figures 5.5a-c is good (see Materials and Methods). Whereas the stability of native WT-apoflavodoxin against unfolding is 10.45 kcal/mol, native Tyr44-apoflavodoxin is severely destabilized compared to WT-apoflavodoxin and has a stability of only 3.29 kcal/mol (Table 5.1). As predicted, substituting Phe44 for Tyr44 affects the stability of the molten globule intermediate only marginally. Unfolding of the Tyr44-intermediate requires 2.11 kcal/mol and 3.74 kcal/mol is needed in case of the WT-folding intermediate. Consequently, maximal population of the intermediate is increased from a mole fraction of 0.63 at 1.76 M GuHCl for WT-apoflavodoxin (Fig. 5.1d) to a mole fraction of 0.86 at 0.84 M denaturant for Tyr44-apoflavodoxin (Fig. 5.5d). In addition, apoflavodoxin's folding intermediate is now detectable in a wider denaturant range. Actually, the unfolding data of Tyr44-apoflavodoxin show that at 25 °C, in absence of GuHCl, already a mole fraction of 0.12 of protein molecules is molten globule.

XY	Y44-apoflavodoxin ΔG (kcal/mol)	WT-apoflavodoxin ΔG (kcal/mol)	Y44-apoflavodoxin m (kcal mol^{-1} M^{-1})	WT-apoflavodoxin m (kcal mol^{-1} M^{-1})
UI	2.11 ± 0.24	3.74 ± 0.49	-1.06 ± 0.07	-1.83 ± 0.19
IN	1.18 ± 0.04	6.70 ± 0.17	-3.73 ± 0.13	-4.4 ± 0.11
UN	3.29 ± 0.24	10.45 ± 0.52		

Table 5.1 Thermodynamic parameters obtained from a three-state global fit of unfolding data of Tyr44-apoflavodoxin (Fig. 5.5) and WT-apoflavodoxin [13] in 100 mM KPPi (see Materials and Methods). ΔG_{XY} is the difference in free energy between species X and Y at 0 M denaturant, and m_{XY} is the dependence of ΔG_{XY} on denaturant concentration. Errors shown are standard errors.

Thermal-induced unfolding experiments confirm that substitution of Phe44 for Tyr44 destabilizes native apoflavodoxin. Using fluorescence emission (Fig. 5.6a), thermal unfolding of only native protein molecules is monitored, as both folding intermediate and unfolded apoflavodoxin have comparable fluorescence signals. Midpoints of unfolding are 48.2 ± 0.1 °C for WT-apoflavodoxin and 32.9 ± 0.3 °C for Tyr44-apoflavodoxin. Using CD data recorded at 210 nm (Fig. 5.6b), it is observed that the thermal midpoint of Tyr44-apoflavodoxin unfolding is 31.0 ± 1.1 °C, which is close to the one detected by fluorescence emission (Fig. 5.6a). The results confirm that at 25 °C, in absence of denaturant, a fraction of Tyr44-apoflavodoxin molecules is present as molten globule.

Upon decreasing KPPi concentration from 100 to 10 mM, Tyr44-apoflavodoxin unfolds and forms the off-pathway molten globule intermediate. Although the substitution of Phe44 for Tyr44 substantially decreases the stability of native protein against unfolding at 25 °C, GuHCl is still required to obtain a maximal population of the molten globule state in 100 mM KPPi, pH 6.0 (Fig. 5.5). Denaturant however, perturbs CD signals at wavelengths below about 210 nm, impeding characterization of the secondary structure content of the folding intermediate of apoflavodoxin.
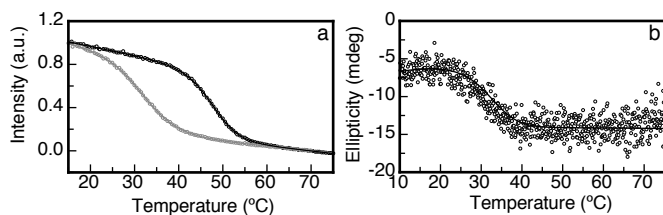
Figure 5.6 Thermal unfolding of apoflavodoxin shows that substitution of Phe44 for Tyr44 destabilizes the protein against thermal unfolding. (a) Thermal unfolding of WT- (black) and Tyr44-apoflavodoxin (gray), as measured by the change in fluorescence emission at 340 nm. (b) Thermal unfolding of Tyr44-apoflavodoxin as measured by CD at 210 nm. This wavelength is chosen because a much larger change in ellipticity is observed than is seen at 222 nm. Sample conditions are 100 mM KPPi, pH 6.0, and the heating rate is 1 °C/min.
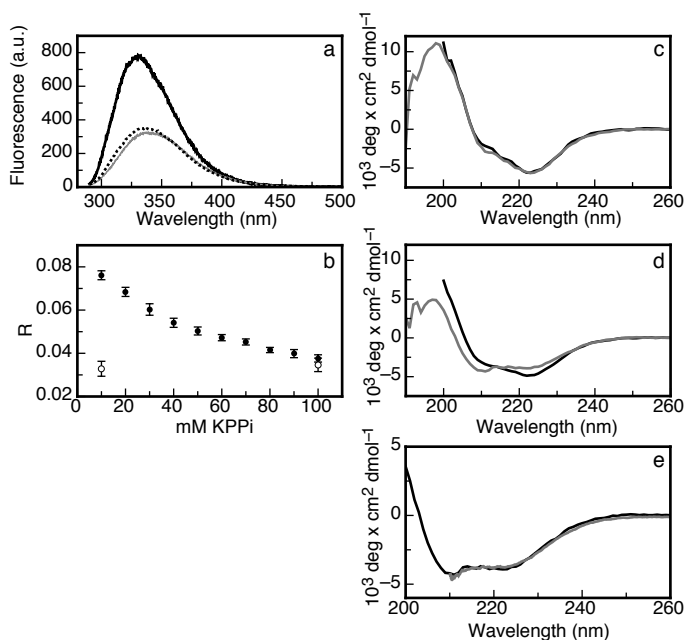

Figure 5.7 Upon decreasing KPPi concentration, Tyr44-apoflavodoxin forms a molten globule. (a) Tryptophan fluorescence spectra of 4 μM Tyr44-apoflavodoxin in 100 mM KPPi (black line), in 10 mM KPPi (dotted line) or in 100 mM KPPi with 0.9 M GuHCl (gray line; at 0.9 M GuHCl mole fractions native, folding intermediate and unfolded molecules are 0.02, 0.86, 0.12, respectively). (b) Fluorescence anisotropy of Tyr44-apoflavodoxin (•) and WT-apoflavodoxin (○), as a function of KPPi concentration. (c, d) Far-UV CD spectra of native WT-apoflavodoxin (c) and Tyr44-apoflavodoxin (d), protein is in 100 mM (black lines) or 10 mM KPPi (gray lines). (e) Far-UV CD spectra of Tyr44-apoflavodoxin in 10 mM KPPi (black line) or in 100 mM KPPi with 0.9 M GuHCl (gray line). pH of all samples is between 6.0 and 6.5, temperature is 25 °C.

Decreasing salt concentration is a well-known manner to destabilize folded proteins [50]. Hence, we reduced buffer concentration in the expectation that the equilibrium between native Tyr44-apoflavodoxin and molten globule shifts towards the latter species. This potential shift in folding equilibrium allows conformational characterization of apoflavodoxin's folding intermediate.

Fluorescence emission of Tyr44-apoflavodoxin changes drastically upon decreasing KPPi concentration from 100 to 10 mM (Fig. 5.7a). Light scattering experiments show that the decrease in intensity is not caused by protein aggregation (data not shown). The fluorescence emission spectrum of Tyr44-apoflavodoxin in 10 mM KPPi strongly resembles the corresponding spectrum of this protein in 100 mM KPPi containing 0.9 M GuHCl. In the latter condition, Tyr44-apoflavodoxin is predominantly present as molten globule with a mole fraction of 0.86, whereas native protein is present with a mole fraction of only 0.02 (Fig. 5.5). As fluorescence emission reports about the microenvironments of tryptophan residues, the spectra obtained suggest that in 10 mM KPPi Tyr44-apoflavodoxin has similar molten globule-like properties as the off-pathway folding intermediate.

Upon decreasing KPPi concentration from 100 to 10 mM, fluorescence anisotropy of WT-apoflavodoxin does not alter within error and remains at the low value typical for natively folded WT-apoflavodoxin molecules (Fig. 5.7b). Consequently, in 10 mM KPPi WT-apoflavodoxin is still native. In contrast, in case of Tyr44-apoflavodoxin, the same decrease in buffer concentration leads to an increase in fluorescence anisotropy from a value of 0.04 to 0.08. Note that upon equilibrium unfolding of native Tyr44-apoflavodoxin to folding intermediate at 0.85 M GuHCl, anisotropy increases from 0.04 to 0.09 (Fig. 5.5c). These observations confirm that in 10 mM KPPi, Tyr44-apoflavodoxin is not native but instead is present as molten globule intermediate.

The far-UV CD spectrum of Tyr44-apoflavodoxin alters considerably upon decreasing KPPi concentration from 100 to 10 mM (Fig. 5.7d). In contrast, in case of WT-apoflavodoxin, no such change is observed (Fig. 5.7c), as the protein remains in its native state. Consequently, a drastic change in secondary structure content of Tyr44-apoflavodoxin occurs upon lowering buffer concentration. The far-UV CD spectrum of Tyr44-apoflavodoxin in 10 mM KPPi strongly resembles the corresponding spectrum of Tyr44-apoflavodoxin in 100 mM KPPi with 0.9 M GuHCl (Fig. 5.7e), the condition at which a mole fraction of 0.86 of Tyr44-apoflavodoxin molecules is molten globule off-pathway folding intermediate.

Taking all observations together, we conclude that upon lowering KPPi concentration from 100 to 10 mM, Tyr44-apoflavodoxin unfolds and forms the off-pathway folding intermediate.

The topology of the molten globule intermediate differs drastically from the one of native apoflavodoxin: it is helical. As the molten globule intermediate can be trapped in absence of denaturant, its conformational characterization is now possible. Native apoflavodoxin contains a parallel β-sheet that is surrounded by five α-helices. It gives rise to a far-UV CD spectrum typical for flavodoxins, i.e., the spectrum has a minimum at 222 nm and a shoulder at 210 nm (Fig. 5.8). In contrast, the far-UV CD spectrum of the Tyr44-folding intermediate under native-like conditions is typical for α-helical proteins. The spectrum has minima at 222 and 210 nm and is similar in shape to the far-UV CD spectrum of the apomyoglobin molten globule at pH 4.3, a species that is helical [51]. Consequently, the topology of apoflavodoxin's molten globule drastically differs from the α-β parallel topology of native apoflavodoxin.
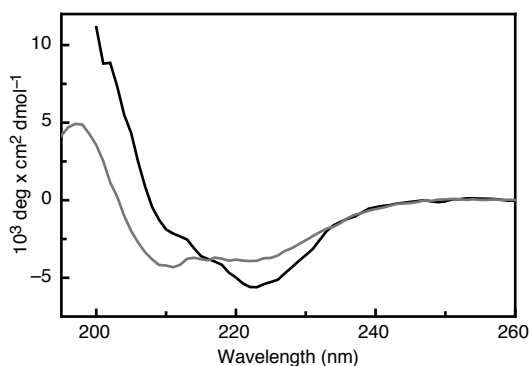
Figure 5.8 The conformation of Tyr44-apoflavodoxin in low salt concentration differs drastically from the conformation of WT-apoflavodoxin in its native state. In low salt concentration, native Tyr44-protein is destabilized and forms a helical molten globule intermediate. Shown are far-UV CD spectra of native WT-apoflavodoxin in 100 mM KPPi (black line) and of Tyr44-apoflavodoxin in 10 mM KPPi (gray line), both at 25 °C. Protein concentration is ~2 µM

A single polypeptide sequence can code for monomeric protein folds that are largely different. In contrast to most folding intermediates studied to date, the molten globule folding species of apoflavodoxin has a remarkably non-native character. What causes the off-pathway folding intermediate of apoflavodoxin to be helical?

Upon protein folding, helices are formed much more rapidly than sheets, especially when parallel β-sheets are involved. This rapid helix formation is due to the highly local character of the interactions in helices. In contrast, the residues that form the strands of a parallel β-sheet are separated by many intervening residues [52, 53], as is the case for apoflavodoxin. Unfolded apoflavodoxin contains four regions with reduced flexibility, of which two transiently form native α-helical structures and one forms non-native α-helical structure [32]. These helices are sufficiently stable to be present about 10% of the time and comprise residues Ala41 – Ile52, Glu104 – Lys118 and Thr160 – Ala169. Rapid formation of α-helices and their subsequent non-native docking through hydrophobic interactions causes formation of the off-pathway intermediate during apoflavodoxin folding [32]. This docking of helices prevents formation of the parallel β-sheet of apoflavodoxin and as a result it seems likely that the intermediate is helical. In the study reported here, conclusive data are presented that show that the molten globule intermediate of apoflavodoxin indeed is of helical nature. Most likely, the regions that form helices in unfolded apoflavodoxin are helical in the off-pathway intermediate as well.

Further support for the absence of a β-sheet in the off-pathway folding intermediate of apoflavodoxin is provided by the observed differences in cooperativity of folding of native apoflavodoxin and molten globule. Native apoflavodoxin is formed highly cooperatively [39], which is inherent to the formation of a parallel β-sheet involving many residues. In contrast, the formation of the off-pathway molten globule of apoflavodoxin is clearly non-cooperative [38]. Consequently, this species cannot contain a β-sheet, as is indeed confirmed by the data presented in this study.

Anfinsen's thermodynamic hypothesis states that each amino acid sequence codes for a single tertiary fold with varying amounts of local flexibility [41]. However, monomeric proteins like for example prions can assume alternative conformations in a multimeric form [28, 29]. The study presented here shows that a single polypeptide sequence can code for monomeric protein folds that are largely different under native-like conditions. The amino acid sequence of apoflavodoxin codes

for the α-β parallel topology of the native state, as well as for a helical protein species. Upon a mild change of conditions, i.e., lowering buffer concentration in case of Tyr44-apoflavodoxin, topological switching between both folds occurs and a monomeric protein species with a distinct fold becomes energetically most favorable. Our observations, together with those reported on lymphotactin [54], show that interconversion between unrelated, monomeric protein structures is likely a common phenomenon in the protein structure universe.

References

1. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45.

2. Jahn, T. R. & Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-117.

3. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-1861.

4. Schuler, B. (2005). Single-molecule fluorescence spectroscopy of protein folding. *Chemphyschem* **6**, 1206-1220.

5. Kay, L. E. (2005). NMR studies of protein structure and dynamics. *J Magn Reson* **173**, 193-207.

6. Schwarzinger, S., Mohana-Borges, R., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2008). Structural characterization of partially folded intermediates of apomyoglobin H64F. *Protein Sci* **17**, 313-321.

7. Whittaker, S. B., Spence, G. R., Gunter Grossmann, J., Radford, S. E. & Moore, G. R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* **366**, 1001-1015.

8. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053-1056.

9. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V. & Fersht, A. R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863-867.

10. Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol* **355**, 139-156.

11. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.

12. Fernandez-Recio, J., Genzor, C. G. & Sancho, J. (2001). Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234-15245.

13. Bollen, Y. J., Sanchez, I. E. & van Mierlo, C. P. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry* **43**, 10475-10489.

14. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X. & Matthews, C. R. (2007). A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol* **366**, 1624-1638.

15. Butler, J. S. & Loh, S. N. (2005). Kinetic partitioning during folding of the p53 DNA binding domain. *J Mol Biol* **350**, 906-918.

16. Kathuria, S. V., Day, I. J., Wallace, L. A. & Matthews, C. R. (2008). Kinetic traps in the folding of βα-repeat proteins: CheY initially misfolds before accessing the native conformation. *J Mol Biol* **382**, 467-484.

17. Otzen, D. E., Giehm, L., Baptista, R. P., Kristensen, S. R., Melo, E. P. & Pedersen, S. (2007). Aggregation as the basis for complex behaviour of cutinase in different denaturants. *Biochim Biophys Acta* **1774**, 323-333.

18. Lorenz, T. & Reinstein, J. (2008). The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP Kinase. *J Mol Biol* **381**, 443-455.

19. Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209-282.

20. Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* **164**, 21-24.

21. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-132.

22. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Adv Protein Chem* **47**, 83-229.

23. Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold Des* **1**, 275-287.

24. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896.

25. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.

26. Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* **365**, 870-880.

27. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-226.

28. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366.

29. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.

30. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem* **280**, 7836-7844.

31. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* **103**, 4095-4100.

32. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

33. van Mierlo, C. P. M., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Sci* **7**, 2331-2344.

34. Visser, N. V., Westphal, A. H., van Hoek, A., van Mierlo, C. P. M., Visser, A. J. & van Amerongen, H. (2008). Tryptophan-tryptophan energy migration as a tool to follow apoflavodoxin folding. *Biophys J* **95**, 2462-2469.

35. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of *Azotobacter vinelandii* holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

36. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

37. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. M. (2008). Extensive formation of off-pathway species during folding of an α-β parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J Am Chem Soc* **130**, 16914-16920.

38. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. (2009). Noncooperative formation of the off-pathway molten globule during folding of the α-β parallel protein apoflavodoxin. *J Am Chem Soc* **131**, 2739-2746.

39. van Mierlo, C. P. M., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

40. Engel, R., Westphal, A. H., Huberts, D. H., Nabuurs, S. M., Lindhoud, S., Visser, A. J. & van Mierlo, C. P. M. (2008). Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem* **283**, 27383-27394.

41. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

42. Steensma, E., Heering, H. A., Hagen, W. R. & Van Mierlo, C. P. M. (1996). Redox properties of wild-type, Cys69Ala, and Cys69Ser *Azotobacter vinelandii* flavodoxin II as measured by cyclic voltammetry and EPR spectroscopy. *Eur J Biochem* **235**, 167-172.

43. Eftink, M. R. (1994). The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys J* **66**, 482-501.

44. Pace, C. N. & Laurents, D. V. (1989). A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-2525.

45. Becktel, W. J. & Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-1877.

46. Privalov, P. L. & Khechinashvili, N. N. (1974). A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J Mol Biol* **86**, 665-684.

47. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

48. Alagaratnam, S., van Pouderoyen, G., Pijning, T., Dijkstra, B. W., Cavazzini, D., Rossi, G. L., Van Dongen, W. M., van Mierlo, C. P. M., van Berkel, W. J. & Canters, G. W. (2005). A crystallographic study of Cys69Ala flavodoxin II from *Azotobacter vinelandii*: structural determinants of redox potential. *Protein Sci.* **14**, 2284-2295.

49. Mayhew, S. G. & Tollin, G. (1992). General properties of flavodoxins. In *Chemistry and biochemistry of flavoenzymes* (Müller, F., ed.), Vol. 3, pp. 389-426. CRC press, Boca Raton, Florida.

50. Timasheff, S. N. & Arakawa, T. (1989). Stabilization of protein structure by solvents. In *Protein Structure: a practical approach* (Creighton, T. E., ed.), pp. 331-345. IRL Press, Oxford.

51. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-1548.
52. Bieri, O. & Kiefhaber, T. (1999). Elementary steps in protein folding. *Biol Chem* **380**, 923-929.
53. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-994.
54. Tuinstra, R. L., Peterson, F. C., Kutlesa, S., Elgin, E. S., Kron, M. A. & Volkman, B. F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc Natl Acad Sci U S A* **105**, 5057-5062.

6

H/D exchange reveals a single stable folding core in apoflavodoxin's molten globule

Sanne M. Nabuurs and Carlo P. M. van Mierlo (submitted)

Abstract

Formation of an off-pathway species during protein folding is typical for proteins with a flavodoxin-like topology. In case of apoflavodoxin from *Azotobacter vinelandii*, this molten globule-like folding intermediate is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap. Full population of this molten globule is possible by replacing Phe44 with Tyr44. In this study, interrupted H/D exchange detected by NMR spectroscopy is used to reveal the stable core of the molten globule of F44Y apoflavodoxin. Exchange rates could be determined for 68 of the 179 backbone amides of this molten globule. Protection factors against exchange show that residues residing in region Leu110 to Val125 have the highest protection factors against exchange and form the stable core of the molten globule. In unfolded apoflavodoxin, in presence of denaturant, four transiently structured regions with restricted flexibility exist and their subsequent non-native docking through hydrophobic interactions causes formation of the folding core of the off-pathway intermediate of apoflavodoxin. The H/D exchange results clearly show that in absence of denaturant one of these four regions, i.e., helical region Leu110 to Val125 is better buried from solvent than the other three ordered regions of apoflavodoxin's molten globule. It is thus tempting to speculate that in the core of apoflavodoxin's molten globule residues Leu110 to Val125 are surrounded by the remaining three ordered regions.

Introduction

Upon folding, protein molecules descend along a funnel describing the free energy of folding, until the folding molecules reach a state that has the lowest free energy, which is the native state [1, 2]. Most folding proteins encounter folding energy landscapes that are rough. As a result, partially folded intermediates, which may be on- or off-pathway to the native state, are populated. When the intermediate is on-pathway, as is observed for the majority of proteins studied to date, it has a native-like topology and is productive for folding. In contrast, when the intermediate is off-pathway, it is trapped in such a manner that the native state cannot be reached without substantial reorganizational events [3]. Several kinetic studies have revealed involvement of off-pathway intermediates during protein folding [4-9].

Characterization of folding intermediates is crucial for the understanding of protein folding, but is challenging due to the transient nature of their existence. The resemblance between early kinetic intermediates and molten globules [10-13] suggests that these molten globules can be considered as models of transient intermediates [3]. This resemblance has been demonstrated for α–lactalbumin [14-16], apomyoglobin [17, 18], RNAse H, [19] T4 lysozyme [20] and Im7 [21]. A decrease of the folding rate due to the presence of an off-pathway molten globule, which is kinetically trapped and partially folded, increases the likelihood of protein aggregation. This aggregation phenomenon can have devastating effects on organisms [22]. Understanding the formation and conformation of molten globules offers potential insights into factors responsible for protein misfolding and, potentially, for numerous debilitating pathologies [23].

Here, we report the characterization of the molten globule folding intermediate of a 179-residue flavodoxin from *A. vinelandii*, using H/D exchange experiments. Flavodoxins are monomeric proteins involved in electron transport, and contain a non-covalently bound FMN cofactor. The proteins consist of a single structural domain and adopt the flavodoxin-like or α-β parallel topology (Fig. 6.1), which is characterized by a parallel β-sheet that is surrounded by α-helices and is widely prevalent in nature. Both denaturant-induced equilibrium and kinetic (un)folding of holoflavodoxin and apoflavodoxin have been characterized using guanidine hydrochloride (GuHCl) as denaturant [6, 24, 25]. The folding data show that apoflavodoxin autonomously folds to its native state, which is structurally identical to holoflavodoxin with exception of residues in the flavin-binding region of the apoprotein. These residues have considerable dynamics in apoflavodoxin [26, 27]. The last step in flavodoxin folding is binding of the cofactor to native apoflavodoxin [25].

Kinetic folding of apoflavodoxin involves an energy landscape with two intermediates and is described by: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows I_{on} \leftrightarrows$ native apoflavodoxin [6]. Intermediate I_{on} lies on the productive route from unfolded to native protein, is highly unstable and is therefore not observed during denaturant-induced equilibrium unfolding. Consequently, GuHCl-induced equilibrium unfolding of apoflavodoxin is described by the three-state model: $I_{off} \leftrightarrows$ unfolded apoflavodoxin $\leftrightarrows$ native apoflavodoxin [6]. Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [6]. The formation of an off-pathway species is typical for proteins with a flavodoxin-like topology [28]. An off-pathway intermediate is experimentally observed for all other α-β parallel proteins of which the kinetic folding has been investigated: apoflavodoxin from *Anabaena* [4], CheY [8], cutinase [5, 7] and UMP/CMP kinase [9].

Spectroscopic data show that the off-pathway species of apoflavodoxin is molten globule-like: it is compact, its three tryptophans are solvent exposed and it has severely broadened NMR resonances

due to exchange between different conformers on the micro- to millisecond timescale [6, 29, 30]. Elevated protein concentrations [30] and molecular crowding [31] cause severe aggregation of this species. Whereas native apoflavodoxin has the flavodoxin-like topology, the molten globule is helical and contains no β-sheet [32]. The presence of this remarkably non-native species shows that a single polypeptide sequence can code for distinct folds that swap upon changing conditions [32].

Apoflavodoxin unfolded in GuHCl has been characterized at the residue-level using heteronuclear NMR spectroscopy [33, 34]. Secondary shifts analysis and investigation of ^{1}H-^{15}N relaxation rates revealed the presence of four structured elements that transiently exist in unfolded apoflavodoxin. These transiently ordered regions have restricted flexibility on the (sub)nanosecond timescale and comprise residues Ala41 – Gly53, Glu72 – Gly83, Gln99 – Ala122 and Thr160 – Gly176 (coloured red in Fig. 6.1) [33]. This restricted flexibility is due to transient helix formation and local and non-local hydrophobic interactions [33, 34]. Upon decreasing denaturant concentration, the four structured elements in unfolded apoflavodoxin transiently interact and subsequently form the ordered core of apoflavodoxin's molten globule [33, 34]. Structure formation within virtually all parts of unfolded apoflavodoxin precedes folding to the molten globule state. This folding transition is non-cooperative and involves a series of distinct transitions [34]. Part of the molten globule remains random coil down to a GuHCl concentration of 1.58 M (i.e., residues Lys13 to Val36, which are coloured blue in Fig. 6.1).

Substitution of Phe44 for Tyr44 leads to severe destabilization of native apoflavodoxin against unfolding and only marginally affects the stability of the molten globule [32]. As a consequence, Tyr44-apoflavodoxin forms the molten globule at low salt concentration, i.e., in 10 mM potassium pyrophosphate (KPPi), whereas in 100 mM KPPi 88% of the protein molecules is native. The molten globule has a low stability against unfolding of approximately 2 kcal/mol [32].

H/D exchange detected by NMR spectroscopy is a powerful technique that can give detailed information about a protein at the level of single amino acids. As Tyr44-apoflavodoxin molecules exist as molten globule at low salt concentration, characterization of an off-pathway intermediate by amide H/D exchange is now possible. In the present work we use interrupted H/D exchange [17, 35, 36] to reveal the stable core of the molten globule of Tyr44-apoflavodoxin in absence of denaturant.

Materials and Methods

Protein Expression and Purification. The single cysteine at position 69 in wild-type *A. vinelandii* (strain ATCC 478) holoflavodoxin II was replaced by an alanine to avoid covalent dimerisation of apoflavodoxin. This protein variant is largely similar to wild-type flavodoxin regarding both redox potential of holoprotein and stability of apoprotein [29, 40]. Subsequently, in addition, the phenylalanine at position 44 was substituted for a tyrosine using site-directed mutagenesis. The latter protein variant is referred to as Tyr44-holoflavodoxin. Uniformly ^{15}N-labeled Tyr44-holoflavodoxin was obtained from transformed *E. coli* cells grown on ^{15}N-labeled algae medium (Silantes, Germany), and purified as described [29].

To obtain native apoprotein, holoprotein was denatured in 6 M GuHCl. Subsequently, both FMN and denaturant were removed via gel filtration on a Superdex 75 column, which is loaded with 100 mM potassium pyrophosphate, pH 6.0. During this removal step, the protein folds to native apoprotein.

H/D exchange. To follow H/D exchange of the molten globule of Tyr44-apoflavodoxin, 1.0 ml samples were prepared of 86 μM native Tyr44-apoflavodoxin in 100 mM potassium pyrophosphate in H_2O, at pH 6.0. Subsequently, the molten globule was populated and H/D exchange was simultaneously initiated by mixing each of these samples with 14 ml of 3.6 mM KPPi in 96% D_2O. The resulting solutions contained 10 mM KPPi, pD 6.72 (uncorrected pH-meter reading). Under these conditions, the molten globule of Tyr44-apoflavodoxin is fully populated, and protein concentration was 5.7 μM. After 10 s to 10 min, H/D exchange was quenched by adding 1 ml of 1.6 mM FMN in 1.45 M KPPi, pH 6.0, in H_2O. The resulting solution initially contained 100 μM of free FMN in 100 mM KPPi, at pD 6.55 (uncorrected pH-meter reading). Under the latter conditions, Tyr44-holoflavodoxin is rapidly formed. The following exchange periods were used: 10 s, 11 s, 20 s, 30 s, 1 min, 2 min, 5 min and 10 min.

To detect the backbone amides that are slowly exchanging in holoflavodoxin, a reference experiment with an exchange period of 10 minutes was performed with the same procedure except that Tyr44-holoflavodoxin instead of Tyr44-apoflavodoxin is used. During this experiment, no molten globule is formed and the protein remains holoprotein throughout the 10 minute H/D exchange period and also during the subsequent steps. This experiment reveals that some H/D exchange is seen to occur for the backbone amides of Asn37 and Leu84 of holoflavodoxin. In total 66 of the 179 backbone amides of the molten globule of apoflavodoxin show no observable exchange in holoflavodoxin during the 10-minute exchange period.

Immediately after H/D exchange, each sample was concentrated by reducing its volume to 500 μL using Amicon centrifugal filter units. Subsequently 50 μL of 1 mM DSS in D_2O was added and each sample was kept frozen at -80 ºC. The NMR samples contained either 91% or 10% D_2O, and pD was 6.55 (uncorrected pH-meter reading). All H/D exchange experiments were performed at 25 ºC.

NMR spectroscopy. Gradient-enhanced ^{1}H-^{15}N HSQC spectra were recorded on a Bruker AMX 500 MHz machine. Sample temperature was 25 ºC. In the 1H dimension, 2048 complex data points were acquired, whereas in the indirect 15N dimension 256 complex data points were collected. With the number of scans set to 64, each experiment lasted 6 hours and 7 minutes.

Spectra were processed and maximal intensities of cross peaks of backbone amides were determined per individual HSQC spectrum. To correct the cross peak intensities for differences in protein concentration between the different NMR samples, after acquisition of each HSQC spectrum 1D proton NMR spectra were acquired (4096 complex data points, with 1024 scans). Subsequently, the integral of the resonance at -1 ppm, which corresponds to a non-exchangeable methyl resonance of holoflavodoxin, was determined. This integral is a direct measure of the protein concentration in the corresponding NMR sample. All 1D proton NMR spectra were acquired three times to estimate the error in the integral mentioned. To obtain relative cross peak intensities, each cross peak intensity is divided by the corresponding integral of the resonance at -1 ppm.
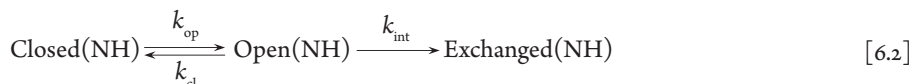
Hydrogen exchange data analysis. A single exponential decay function was fitted to the time-dependent relative cross peak intensities:

$$I(t) = I(\infty) + C \cdot \exp(-k_{ex} \cdot t) \tag{6.1}$$

In Equation 6.1, t is the time between initiation of H/D exchange and quenching of this exchange, $I(\infty)$ is the peak intensity at infinite time, C is the pre-exponential factor, and k_{ex} is the amide proton exchange rate.

Peak intensities at infinite time, $I(\infty)$, are above zero for all backbone amides. For some residues $I(\infty)$ equals the maximum noise level in the corresponding cross peak area of the HSQC spectrum. For other residues $I(\infty)$ is larger than the noise level dus to presence of 9% H_2O causing incomplete exchange of the corresponding backbone amides with D_2O.

Model for H/D exchange. Quantitative interpretation of H/D exchange is possible using a simple model [41]:

$$\text{Closed(NH)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open(NH)} \xrightarrow{k_{int}} \text{Exchanged(NH)} \qquad [6.2]$$

In this model, the open or exchange-competent form and the closed or exchange-incompetent form of a protein at the site of a particular amide proton interconvert with rate constants for opening (k_{op}) and closing (k_{cl}). From the open state, exchange takes place with the intrinsic rate constant k_{int}. The time course of exchange is monitored at the residue level by NMR spectroscopy, since amide protons give rise to a ^{1}H-NMR signal, and replacement of the proton by a deuteron leads to disappearance of this signal. Under conditions favoring the closed state (i.e. $k_{op} \ll k_{cl}$), the observed exchange rate k_{ex} equals:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{cl} + k_{int}} \qquad [6.3]$$

Depending on the ratio of k_{cl} and k_{int}, two limiting cases may be reached. If $k_{cl} \ll k_{int}$ Equation 6.3 reduces to:

$$k_{ex} = k_{op} \qquad [6.4]$$

Under this so-called EX1 condition k_{ex} of a certain amide proton informs about the rate constant for local conversion between the closed and the open state of its micro-environment. If $k_{cl} \gg k_{int}$ Equation 6.3 reduces to:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{cl}} = K_{op}k_{int} \qquad [6.5]$$

Under this so-called EX2 condition, the ratio of k_{ex} and k_{int} provides the equilibrium constant K_{op} for local opening of the protein structure.

Results and Discussion

To probe hydrogen exchange rates of protein species that are marginally stable and thus have rapidly exchanging amide protons, as is the case for the molten globule of Tyr44-apoflavodoxin, amide proton exchange needs to be interrupted. In the interrupted H/D exchange methodology [17, 35, 36], an unstable protein species of interest is kept in D_2O for a variable time period, t, during which exchange of amide protons for solvent deuterons occurs. After this period, conditions are changed

such that the protein arrives into circumstances in which no, or very slow, hydrogen exchange occurs and hence exchange is effectively quenched. Now, by using NMR spectroscopy, exchange rates, k_{ex}, of amide protons in the species of interest can be inferred.

Interrupted H/D exchange tracks 68 backbone amides of the molten globule of Tyr44-apoflavodoxin. Induction of the molten globule folding intermediate of Tyr44-apoflavodoxin and simultaneous initiation of H/D exchange is achieved in the following manner. A solution that contains native Tyr44-apoflavodoxin and 100 mM KPPi in water is diluted tenfold through addition of an appropriate volume of D_2O. This procedure leads to a 10 mM KPPi solution containing the molten globule of Tyr44-apoflavodoxin in D_2O. In 10 mM KPPi, the molten globule is fully populated within the dead time of the experiment (data not shown). After time periods, t, of 10 s to 10 min, H/D exchange is quenched by increasing the salt concentration and by concomitant addition of an excess of FMN. This quenching leads to rapid reconstitution of holoflavodoxin (see Materials and methods), in which exchange of many amide protons is very slow (half-times of exchange, $1/k_{ex}$, of more than 100 days) [27] (Bollen and van Mierlo, personal communication). Subsequently, ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of the different holoflavodoxin samples are recorded (Fig. 6.2). The series of HSQC spectra, as a function of exchange time, t, allows the determination of the exchange rates of backbone amides of the molten globule of Tyr44-apoflavodoxin.

A prerequisite for successful use of the interrupted exchange experiment is that native holoflavodoxin is obtained after quenching of H/D exchange. To verify the latter, in a reference experiment the interrupted H/D exchange methodology described above is used, except that in all
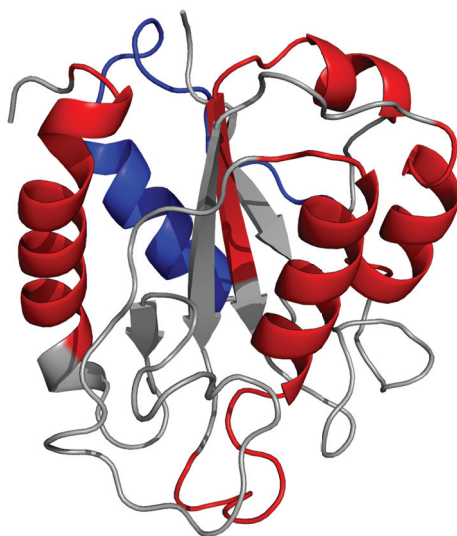

Figure 6.1 Cartoon drawing of native flavodoxin from *A. vinelandii* (pdb ID 1YOB [42]). The FMN cofactor is not shown. The cartoon drawing shows in red regions of unfolded apoflavodoxin that have restricted flexibility on the (sub)nanosecond timescale [33]. The region of apoflavodoxin's molten globule that remains random coil down to a GuHCl concentration of 1.58 M is highlighted in blue [34].
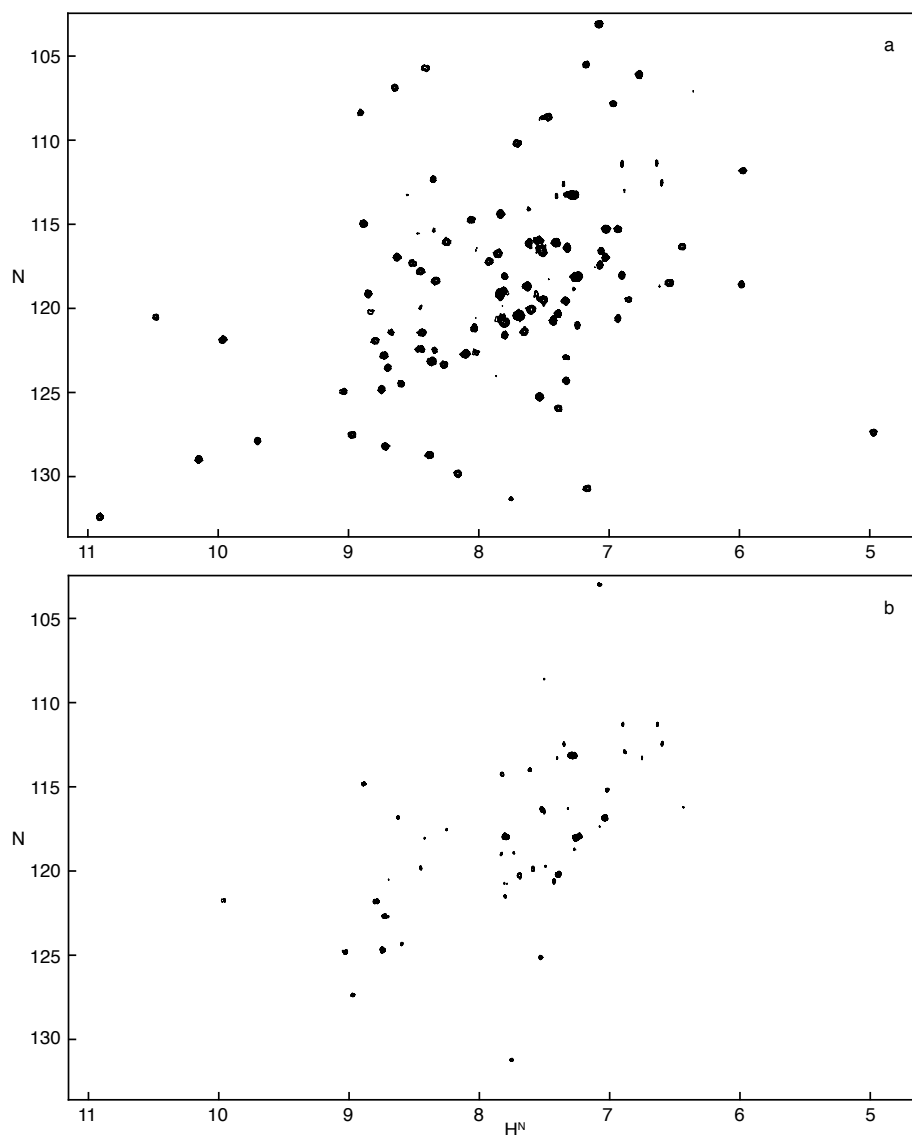
Figure 6.2 Demonstration of the effects of H/D exchange of backbone amides of the molten globule of Tyr44-apoflavodoxin on the corresponding ¹H-¹⁵N HSQC spectrum of holoflavodoxin. (a) HSQC spectrum of holoflavodoxin in D₂O. (b) HSQC spectrum of the holoflavodoxin sample obtained from an interrupted H/D exchange experiment during which the molten globule of Tyr44-apoflavodoxin was allowed to exchange its amide protons for solvent deuterons for a period of 2 minutes (see Materials and methods).

steps H_2O instead of D_2O is used and thus no exchange occurs. This experiment indeed demonstrates that properly folded holoflavodoxin is obtained, as the corresponding HSQC spectrum is indistinguishable from the one of freshly purified Tyr44-holoflavodoxin (data not shown).

The only amide probes capable of reporting on properties of the molten globule of Tyr44-apoflavodoxin in the interrupted H/D exchange experiment are those that are slowly exchanging in holoflavodoxin. In addition, cross peaks of these amides in HSQC spectra of holoprotein should not suffer from overlap. Due to these limitations, time-dependent amide proton exchange curves and corresponding exchange rates could be determined for 68 of the 179 backbone amides of the molten globule of Tyr44-apoflavodoxin (Table 6.1, and Materials and methods). Examples of four typical exchange curves are shown in Figure 6.3.

Most backbone amides of the molten globule of Tyr44-apoflavodoxin fully exchange within 10 minutes (Figs. 6.3a,b). Partial exchange occurs for the backbone amides of Leu50 (Fig. 6.3c), Ile51, Val91, Val100, Gly111, Phe116, Lys118 and Val125. The backbone amide of Tyr114 does not exchange at all within 10 minutes and its average maximal intensity in this time period is 2.3 ± 0.3 (Fig. 6.3d). Similar maximal intensities, extrapolated to an exchange period of zero seconds, are observed for the HSQC cross peaks of the other detectable backbone amides.

Exchange rates and protection factors. Half times of exchange of the backbone amides of the molten globule of Tyr44-apoflavodoxin are shown in Figure 6.4a. Due to the low stability against unfolding of the molten globule, the corresponding observable amide protons exchange much faster than those of holoflavodoxin. Half times of exchange of the molten globule do not exceed 500 s (except for the backbone amide of Tyr114).

Because interconversion between unfolded apoflavodoxin and the molten globule is fast (rate constant exceeds 400 s^{-1}, [6]) exchange is expected to occur according to an EX2 mechanism of exchange (see Materials and methods). Consequently, protection factors against exchange can be calculated by dividing the intrinsic amide proton exchange rates, k_{int}, by the measured amide proton exchange rates. The k_{int} values are determined by using free peptide exchange rates, which are
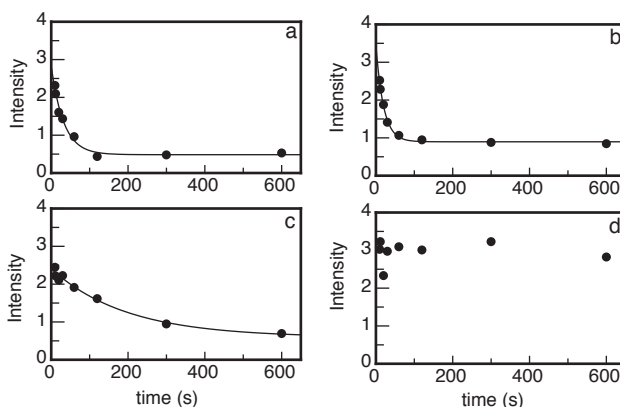

Figure 6.3 Typical H/D exchange curves obtained for the molten globule-like folding intermediate of Tyr44-apoflavodoxin. Shown are the time-dependent decreases of the maximal intensities of cross peaks arising from the backbone amides of Ile-3 (k_{ex} is 0.031 ± 0.009 s^{-1}) (a), Ala-18 (k_{ex} is 0.050 ± 0.016 s^{-1}) (b), Leu-50 (k_{ex} is 0.0051 ± 0.0023 s^{-1}) (c), and Tyr-114 (k_{ex} is not determined) (d), respectively. Data are extracted from ^{1}H-^{15}N HSQC spectra like the ones shown in Figure 6.2.

Table 6.1 H/D exchange data of backbone amides of the molten globule of Tyr44-apoflavodoxin in 10 mM KPPi, pD 6.72, at 25 °C. Data are obtained from interrupted H/D exchange experiments. Shown are rates of H/D exchange, k_{ex}, and corresponding standard deviations, as well as protection factors, PF, and corresponding standard deviations.

Residue	k_{ex} s^{-1}	sd k_{ex} s^{-1}	PF	sd PF
Ile 3	3.07e-2	9.20e-3	7.04	2.11
Gly 4	6.88e-2	2.75e-2	14.03	5.6
Leu 5	2.83e-2	9.16e-3	12.12	3.93
Phe 6	2.02e-2	6.68e-3	15.44	5.1
Phe 7	2.68e-2	8.42e-3	21.67	6.8
Gly 8	8.06e-2	4.64e-2	23.35	13.44
Thr 14	≥ 0.5 [a]			
Arg 15	≥ 0.5 [a]			
Lys 16	7.03e-2	2.80e-2	18.96	7.55
Val 17	9.84e-2	3.19e-2	2.35	0.76
Ala 18	5.01e-2	1.62e-2	12.74	4.12
Lys 19	1.08e-1	4.01e-2	7.42	2.75
Ser 20	6.48e-2	2.28e-2	41.96	14.74
Ile 21	8.53e-2	3.34e-2	3.84	1.5
Lys 22	1.10e-1	5.84e-2	4.31	2.3
Lys 23	4.79e-2	1.13e-2	22.08	5.19
Phe 25	5.22e-2	1.61e-2	16.11	4.98
Met 30	5.28e-2	3.54e-2	25.83	17.32
Leu 34	4.07e-2	3.08e-2	5.68	4.3
Asn 37	1.87e-1	1.56e-1	10.52	8.77
Phe 49	9.21e-3	2.24e-3	87.22	21.19
Leu 50	5.10e-3	2.30e-3	52.18	23.51
Ile 51	2.47e-3	1.80e-3	40.91	29.84
Leu 52	7.75e-3	3.12e-3	17.6	7.09
Gly 53	2.44e-2	1.12e-2	41.37	18.97
Thr 54	4.77e-2	1.40e-2	23.25	6.81
Glu 76	4.23e-2	6.43e-2	4.55	6.91
Phe 77	1.45e-2	4.29e-3	24.68	7.29
Leu 78	5.59e-3	1.90e-3	47.54	16.14
Ile 81	1.33e-2	4.58e-3	16.22	5.57
Glu 82	4.62e-2	1.29e-2	3.47	0.97
Leu 84	2.37e-2	1.56e-2	14.47	9.54
Val 91	4.13e-3	1.64e-3	67.42	26.84
Ala 92	1.48e-2	4.16e-3	43.03	12.09
Leu 93	9.63e-3	3.15e-3	24.03	7.85
Phe 94	8.63e-3	2.57e-3	36.19	10.76
Gly 95	5.67e-2	1.80e-2	33.19	10.53
Leu 96	2.34e-2	1.25e-2	14.64	7.8
Asp 98	≥ 0.5 [a]			
Val 100	2.32e-2	1.21e-2	11.98	6.24
Gly 101	1.71e-1	1.65e-1	6.95	6.69
Tyr 102	4.22e-2	2.06e-2	16.58	8.08
Tyr 106	5.03e-2	1.60e-2	19.62	6.25
Asp 108	3.19e-2	7.35e-3	8.54	1.97
Leu 110	2.73e-3	1.22e-3	84.83	37.99
Gly 111	7.64e-3	2.36e-3	132.22	40.87
Ser 115	1.29e-2	2.27e-3	179.64	31.61
Phe 116	3.47e-3	1.37e-3	291.64	115.24
Lys 118	2.97e-3	1.16e-3	310.19	121.12
Arg 120	1.06e-2	2.82e-3	66.09	17.6
Ala 122	1.17e-2	1.79e-3	111.1	16.98
Lys 123	1.55e-2	4.14e-3	51.9	13.88
Val 125	5.14e-3	1.53e-3	20.12	6
Trp 128	6.17e-2	2.93e-2	11.07	5.25
Ala 140	5.99e-2	1.95e-2	10.4	3.38
Val 141	4.53e-2	1.04e-2	3.88	0.89
Val 142	3.46e-2	1.02e-2	3.68	1.09
Phe 146	7.01e-2	3.36e-2	9.53	4.57
Val 147	2.64e-2	1.16e-2	7.65	3.38
Gly 148	8.12e-2	2.56e-2	14.62	4.61
Ala 150	3.87e-2	1.68e-2	14.02	6.1
Leu 151	1.66e-2	6.20e-3	13.96	5.21
Gln 156	4.85e-2	1.98e-2	43.51	17.78
Arg 163	3.23e-2	1.34e-2	23.16	9.59
Val 164	2.61e-2	5.75e-3	11.18	2.47
Trp 167	1.54e-2	3.45e-3	22.27	4.99
Leu 168	7.72e-3	1.75e-3	23.27	5.26
Gln 170	2.05e-2	3.77e-3	49.31	9.07
Ile 171	1.50e-2	3.79e-3	17.37	4.4
Ala 172	1.26e-2	2.20e-3	41.24	7.21
Phe 175	4.72e-2	1.07e-2	7.59	1.72

[a] A cross peak is observed for the corresponding backbone amide in a reference H/D exchange experiment during which the protein remains holoflavodoxin throughout a 10 minute exchange period (see Materials and methods). However, in a regular interrupted H/D exchange experiment using the shortest achievable exchange period of 10 s, a period during which the protein is molten globule, no cross peak is observed for this backbone amide. Consequently, the amide must have exchanged within 10 s, and the corresponding kex-value must be ≥ 0.5 s^{-1}.
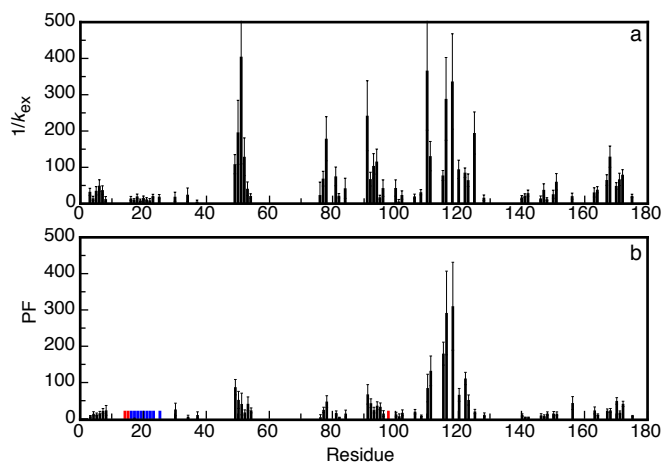
Figure 6.4 Protection against H/D exchange of the backbone amides of the molten globule of Tyr44-apoflavodoxin. (a) Half-times of exchange, $1/k_{ex}$, could be determined for 68 residues. The corresponding protection factors, PF, are shown in (b). Residues with backbone amide protons that fully exchange within the 10-second dead time of the interrupted H/D exchange experiment are indicated in red. Error bars shown are standard deviations. No protection factor can be determined for residues Lys16 to Phe25 (indicated in blue), as the corresponding backbone amides exchange according to an EX1 mechanism.
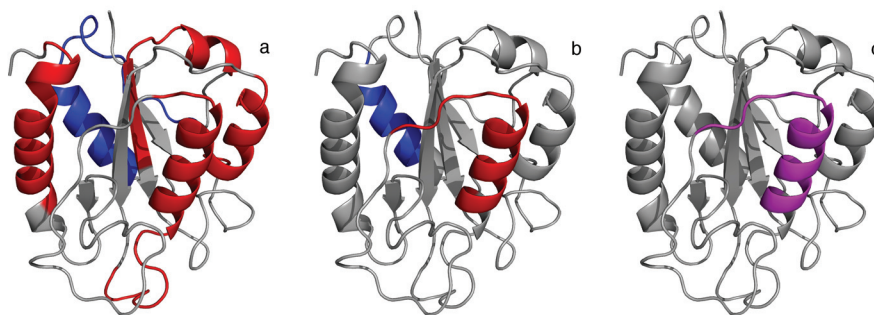

Figure 6.5 Overview of several folding properties of apoflavodoxin using cartoon drawings of native flavodoxin from *A. vinelandii* (pdb ID 1YOB). (a) The four regions of unfolded apoflavodoxin with restricted flexibility on the (sub)nanosecond timescale [33] are colored red; assembly of these structured elements leads to formation of the ordered core of apoflavodoxin's molten globule [34]. The region of apoflavodoxin's molten globule that remains random coil down to a GuHCl concentration of 1.58 M [32] is colored blue. (b) Residues Leu110 to Val125 of the molten globule of Tyr44-apoflavodoxin have the highest protection factors against H/D exchange and are colored red; the region of the molten globule that probably exchanges its backbone amides via an EX1 mechanism is colored blue. In unfolded apoflavodoxin in 3.4 M GuHCl residues Glu104 to Lys118 transiently form an α-helix [33]. (c) Partially unfolded form PUF4 of apoflavodoxin as detected by native-state H/D exchange [24]. In PUF4 only the amides of the residues that are coloured purple are protected against H/D exchange and all other parts of the protein are unfolded.
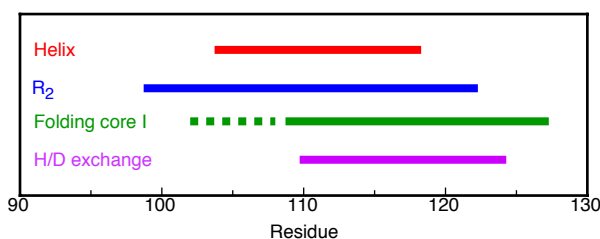
Figure 6.6 Overview of experimental data obtained for residues 99 – 127 of the molten globule of apoflavodoxin. Residues Glu104 to Lys118 of unfolded apoflavodoxin in 3.4 M GuHCl are coloured red, as these residues transiently form an α-helix [33]. Residues Gln99 to Ala122 of unfolded apoflavodoxin in 3.4 M GuHCl are coloured blue, as transverse ^{1}H-^{15}N relaxation rates show that these residues have restricted flexibility on the (sub)nanosecond timescale [33]. Residues Ala109 to Ser127 are coloured green, as they belong to the group of residues that have the largest midpoints of folding towards the molten globule (i.e., 2.64 ± 0.05 M GuHCl) and belong to its ordered core [34]. Unfortunately, no midpoints of unfolding could be determined for residues that reside in the dashed region. Residues Leu110 to Val125 of the molten globule of Tyr44-apoflavodoxin are coloured purple, as these residues have the highest protection factors against H/D exchange.

corrected for the effects of local amino acid sequence and calibrated for the pH and temperature of the exchange experiment according to Bai *et al* [37]. Protection factors against exchange of the molten globule are shown in Figure 6.4b. With exception of Tyr114, these protection factors are low and do not exceed a value of approximately 300 (Table 6.1). These protection factors are comparable to those observed for the molten globules of apomyoglobin [36], apoleghemoglobin [38] and RNAse HI [39]. In contrast, backbone amides of native apoflavodoxin have protection factors up to 2×10^6 (Bollen and van Mierlo, personal communication).

Remarkably, the half times of exchange of residues Lys16 to Phe25 are equal within error ($1/k_{ex}$ = 14.4 ± 4.7 s, Fig. 6.4a). This equality suggests that the corresponding backbone amides become exposed to solvent in a cooperative unfolding event and that exchange occurs before these residues refold to a structured element that is protected against exchange (i.e., exchange occurs according to an EX1 mechanism of exchange, see Materials and methods). Indeed, we previously showed that Lys16 to Phe25 belong to the most unstable part of apoflavodoxin's molten globule, as these residues remain random coil down to a GuHCl concentration of 1.58 M [34] (Figs. 6.5a,b). As a result, no protection factor can be determined for residues Lys16 to Phe25.

Residues residing in region Leu110 to Val125 of the molten globule have the highest protection factors (Figs. 6.4b and 6.5b). In this region, five residues, i.e., Gly111, Ser115, Phe116, Lys118 and Ala122, have protection factors that exceed a value of 100. In case of the backbone amide of Tyr114 the protection factor could not be determined, since this amide does not exchange at all during a 10-minute exchange period (Fig. 6.3d). Consequently, the corresponding protection factor must be well above 300. Using heteronuclear NMR spectroscopy and GuHCl-induced unfolding, residues Leu110 to Val125 have previously been identified to be part of the ordered core of apoflavodoxin's molten globule. [34] In addition, these residues have restricted flexibility in unfolded apoflavodoxin [33] (Fig. 6.5a). Although residues Phe49 to Thr54, Phe77 to Leu84, Val91 to Gly95, and Thr160 to Leu179 have also been identified to be part of this ordered core [34], the corresponding protection factors show that these residues are much less protected against exchange.

The molten globule folding intermediate of apoflavodoxin is almost entirely structured. Upon going from native protein to molten globule the hydrodynamic radius of apoflavodoxin expands only by about 11%, as demonstrated by FCS experiments [31]. The molten globule is thus relatively compact. Far-UV CD reports that the molten globule contains helices [32]. Fluorescence anisotropy of the tryptophans of the molten globule equals a value of 0.08, whereas fluorescence anisotropy of unfolded apoflavodoxin in GuHCl is 0.05 [32]. Thus, in the molten globule of apoflavodoxin its three tryptophans are immobilised. Chemical shift changes of the NMR resonances of unfolded apoflavodoxin upon decreasing GuHCl concentration to 1.58 M show that structure formation in virtually all parts of the unfolded protein precedes folding to the molten globule state. The molten globule is ordered and only residues 21 to 36 behave as a random coil at 1.58 M GuHCl [34].

The H/D data presented in this study reveal that in absence of denaturant apoflavodoxin's molten globule must be almost entirely structured. Rates of backbone amide exchange could be determined for 68 of its residues, which are spread along its sequence (Fig. 6.4a). No exchange rates could be determined for the remaining residues, mainly because these residues exchange relatively fast in holoflavodoxin (Bollen and van Mierlo, personal communication). The parts of the molten globule for which exchange rates are determined cannot be random coil, as in that situation exchange of the corresponding backbone amides would be completed during the 10 seconds lasting dead time of the interrupted H/D exchange experiment (see Materials and methods) and no exchange curves of these residues would be obtained. The data of Figure 6.4a show that in absence of denaturant also several of the residues that reside in the region comprising residues 21 to 36 of the molten globule are protected against exchange. Thus, the corresponding region of apoflavodoxin's molten globule cannot be random coil in absence of denaturant but is structured instead.

Residues Leu110 to Val125 form the stable core of the off-pathway molten globule of apoflavodoxin. Apoflavodoxin unfolded in 3.4 M GuHCl contains four regions with restricted flexibility (Fig. 6.5a). [33, 34] The residues of apoflavodoxin's molten globule that have the highest midpoints against unfolding by GuHCl roughly coincide with those residues that are transiently ordered in unfolded apoflavodoxin. Upon decreasing denaturant concentration, assembly of these four structured elements in unfolded apoflavodoxin through hydrophobic interactions leads to formation of the ordered core of apoflavodoxin's off-pathway molten globule [34].

The wealth of information obtained until now for residues 99 to 127 of apoflavodoxin, including the results obtained in the study presented here, is summarized in Figure 6.6. In unfolded apoflavodoxin residues Glu104 to Lys118 transiently form an α-helix [33]. In addition, restricted flexibility on the (sub)nanosecond timescale has been observed in unfolded apoflavodoxin for residues Gln99 to Ala122 [33]. Midpoints of unfolding, as determined at the residue level by NMR spectroscopy, show that part of the ordered core of the off-pathway molten globule is formed by residues Ala109 to Ser127 [34]. The H/D exchange data presented here show that residues Leu110 to Val125 of the molten globule have the highest protection factors against exchange. In addition, in a previous study, detection by NMR spectroscopy of native state H/D exchange in presence of small amounts of a denaturant identified five unfolding clusters of residues within native apoflavodoxin [24]. Four of these clusters unfold subglobally in a cooperative manner. The resulting conformations are partially unfolded forms (PUFs) of the protein. Both PUF1 and PUF2 are unfolding excursions that start from native apoflavodoxin but do not continue to the unfolded state and do not reside on the productive folding route. In contrast, both PUF3 and PUF4 probably are PUFs of the off-pathway folding intermediate [24]. In PUF4 only the backbone amides of residues Leu110 to Val125 are

inaccessible to water (Fig. 6.5c).

Comparison of Figures 6.5b and 6.5c shows that the residues of the molten globule of Tyr44-apoflavodoxin that are protected most against H/D exchange coincide with those that are inaccessible to water in PUF4. Consequently, now conclusive proof is obtained that PUF4 indeed is an unfolding excursion of apoflavodoxin's off-pathway intermediate. The collection of data obtained implies that the residues that are protected against exchange in PUF4 are helical.

Three transiently formed helices, including the one formed by residues Glu104 to Lys118, are equally populated for about 10% of the time in unfolded apoflavodoxin in 3.4 M GuHCl [33]. Consequently, in unfolded apoflavodoxin these helices are equally stable against unfolding and thus are equally protected against H/D exchange. Upon decreasing denaturant concentration, assembly of these helices leads to formation of the ordered core of apoflavodoxin's molten globule [34]. This molten globule is remarkably helical and contains no β-sheet [32]. Further insight into the conformational properties of this molten globule is obtained from the H/D exchange data presented here. These data show that one of the helices of the off-pathway molten globule (i.e., the region comprising residues Leu110 to Val125; Figs. 6.4b and 6.5b) is better protected against exchange then the others. The enhanced protection against exchange suggests that this helix is better buried in apoflavodoxin's molten globule compared to the other helices mentioned. Hydrophobic interactions of this helix with the other ordered parts of the molten globule, although loose in nature, cause context-dependent stabilisation of this helix against unfolding. Most likely, in apoflavodoxin's molten globule the ordered regions detected in unfolded apoflavodoxin surround residues Leu110 to Val125. Residues Leu110 to Val125 thus form the stable core of the helical molten globule of the α-β parallel protein apoflavodoxin. Non-native docking of helices in apoflavodoxin's molten globule prevents formation of the parallel β-sheet of native apoflavodoxin. Hence, to produce native α-β parallel protein molecules, the off-pathway species needs to unfold.

References

1.   Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.
2.   Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.
3.   Jahn, T. R. & Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-117.
4.   Fernandez-Recio, J., Genzor, C. G. & Sancho, J. (2001). Apoflavodoxin folding mechanism: an alpha/beta protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234-15245.
5.   Melo, E. P., Chen, L., Cabral, J. M., Fojan, P., Petersen, S. B. & Otzen, D. E. (2003). Trehalose favors a cutinase compact intermediate off-folding pathway. *Biochemistry* **42**, 7611-7617.
6.   Bollen, Y. J., Sanchéz, I. E. & van Mierlo, C. P. M. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry* **43**, 10475-10489.
7.   Otzen, D. E., Giehm, L., Baptista, R. P., Kristensen, S. R., Melo, E. P. & Pedersen, S. (2007). Aggregation as the basis for complex behaviour of cutinase in different denaturants. *Biochim Biophys Acta* **1774**, 323-333.
8.   Kathuria, S. V., Day, I. J., Wallace, L. A. & Matthews, C. R. (2008). Kinetic traps in the folding of beta alpha-repeat proteins: CheY initially misfolds before accessing the native conformation. *J Mol Biol* **382**, 467-484.
9.   Lorenz, T. & Reinstein, J. (2008). The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP Kinase. *J Mol Biol* **381**, 443-455.
10. Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* **164**, 21-24.
11. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Adv Protein Chem* **47**, 83-229.
12. Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209-282.
13. Redfield, C. (2004). Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34**, 121-132.
14. Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold Des* **1**, 275-287.
15. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. & Dobson, C. M. (1999). Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J Mol Biol* **288**, 673-688.
16. Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K., Kihara, H., Amemiya, Y. & Kuwajima, K. (2002). Fast compaction of alpha-lactalbumin during folding studied by stopped-flow X-ray scattering. *J Mol Biol* **321**, 121-132.
17. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-1548.
18. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896.
19. Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* **4**, 298-304.
20. Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J Mol Biol* **365**, 870-880.
21. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-226.
22. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366.
23. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.
24. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci USA* **103**, 4095-4100.
25. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem* **280**, 7836-7844.
26. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

27. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. (1998). Apparent local stability of the secondary structure of *Azotobacter vinelandii* holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

28. Bollen, Y. J. & van Mierlo, C. P. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

29. van Mierlo, C. P., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Sci* **7**, 2331-2344.

30. van Mierlo, C. P., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

31. Engel, R., Westphal, A. H., Huberts, D. H., Nabuurs, S. M., Lindhoud, S., Visser, A. J. & van Mierlo, C. P. M. (2008). Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem* **283**, 27383-27394.

32. Nabuurs, S. M., Westphal, A. H., aan den Toorn, M., Lindhoud, S. & van Mierlo, C. P. M. (2009). Topological switching between an α-β parallel protein and a remarkably helical molten globule. (*submitted*).

33. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. (2008). Extensive formation of off-pathway species during folding of an alpha-beta parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J Am Chem Soc* **130**, 16914-16920.

34. Nabuurs, S. M., Westphal, A. H. & van Mierlo, C. P. M. (2009). Non-cooperative formation of the off-pathway molten globule during folding of the α-β parallel protein apoflavodoxin. *J Am Chem Soc* **131**, 2739-2746.

35. Engel, M. F., Visser, A. J. & Van Mierlo, C. P. M. (2004). Conformation and orientation of a protein folding intermediate trapped by adsorption. *Proc Natl Acad Sci USA* **101**, 11316-11321.

36. Nishimura, C., Dyson, H. J. & Wright, P. E. (2005). Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc Natl Acad Sci USA* **102**, 4765-4770.

37. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.

38. Nishimura, C., Dyson, H. J. & Wright, P. E. (2008). The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. *J Mol Biol* **378**, 715-725.

39. Yamasaki, K., Yamasaki, T., Kanaya, S. & Oobatake, M. (2003). Acid-induced denaturation of *Escherichia coli* ribonuclease HI analyzed by CD and NMR spectroscopies. *Biopolymers* **69**, 176-188.

40. Steensma, E., Heering, H. A., Hagen, W. R. & Van Mierlo, C. P. M. (1996). Redox properties of wild-type, Cys69Ala, and Cys69Ser *Azotobacter vinelandii* flavodoxin II as measured by cyclic voltammetry and EPR spectroscopy. *Eur J Biochem* **235**, 167-172.

41. Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Adv Protein Chem* **21**, 287-386.

42. Alagaratnam, S., van Pouderoyen, G., Pijning, T., Dijkstra, B. W., Cavazzini, D., Rossi, G. L., Van Dongen, W. M., van Mierlo, C. P. M., van Berkel, W. J. & Canters, G. W. (2005). A crystallographic study of Cys69Ala flavodoxin II from *Azotobacter vinelandii*: structural determinants of redox potential. *Protein Sci* **14**, 2284-2295.

7

Summarizing discussion

Summary

The question: "Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?" is the main subject of this thesis. *A. vinelandii* apoflavodoxin is chosen as protein of interest as it is a representative of α-β parallel proteins, which are widely prevalent in nature. The folding behavior of *A. vinelandii* apo- and holoflavodoxin has been studied extensively during the past years. Both denaturant-induced equilibrium and kinetic (un)folding of apoflavodoxin have been characterized in detail using GuHCl as denaturant [1-8]. An off-pathway intermediate plays a major role during apoflavodoxin folding and is also observed during the kinetic folding of other proteins with an α-β parallel topology of which the folding mechanism has been studied [9].

Approximately 90% of folding molecules fold via off-pathway intermediate I_{off}, which is a relatively stable species that needs to unfold to produce native protein and thus acts as a trap [3]. Residual structure in the unfolded state of apoflavodoxin probably facilitates formation of this species. In **chapter 2** detailed information about unfolded apoflavodoxin is revealed by heteronuclear NMR spectroscopy. In 6.0 M GuHCl apoflavodoxin behaves as a random coil as is shown by far-UV CD and by ^{1}H-^{15}N R_2 relaxation rates. Upon lowering denaturant concentration the amount of residual structure in apoflavodoxin increases. Chemical shift deviations between unfolded apoflavodoxin in 3.4 and 6.0 M GuHCl reveal in unfolded apoflavodoxin in 3.4 M GuHCl the presence of three transiently formed α-helices and of one structured region that is neither an α-helix nor a β-sheet. One of these transiently formed α-helices is non-native, and a part of this helix becomes a β-strand in native apoflavodoxin. Four regions with restricted flexibility on the (sub)nanosecond time scale are revealed by ^{1}H-^{15}N R_2 relaxation rates of unfolded apoflavodoxin in 3.4 M GuHCl. These four regions coincide with the ordered regions found by chemical shift analysis and match with regions of large AABUF (average area buried upon folding), which is correlated with hydrophobicity [10]. Chemical shift deviations upon substitution of a glutamine residue with a more hydrophobic cysteine residue on position 48, in the middle of the non-native α-helix in unfolded apoflavodoxin, show that this non-native helix has hydrophobic interactions with all other ordered regions in unfolded apoflavodoxin. Formation of native and non-native helices in unfolded apoflavodoxin and subsequent docking of these helices leads to formation of a compact off-pathway intermediate.

The formation of this off-pathway intermediate is discussed in **chapter 3**. Backbone amide resonances of unfolded apoflavodoxin are followed in a series of ^{1}H-^{15}N HSQC spectra acquired at concentrations of GuHCl between 4.05 M and 1.58 M. Analysis of cross peak disappearance of unfolded backbone amides made it possible to determine midpoints of unfolding of 68 backbone amides. Residues were grouped in five different groups according to their midpoint of unfolding. The group with the highest C_m value forms the folding core of the molten globule of apoflavodoxin in presence of GuHCl. This folding core roughly coincides with the regions with restricted flexibility in unfolded apoflavodoxin. The core is gradually extended upon decreasing denaturant concentration, but part of apoflavodoxin's molten globule remains random coil in the denaturant range investigated. The formation of the off-pathway intermediate of apoflavodoxin is non-cooperative and involves a series of distinct transitions in contrast to the cooperative formation of native apoflavodoxin [7]. In addition, chemical shifts of the amides of unfolded apoflavodoxin could be tracked over the denaturant range investigated. Analysis of the chemical shift changes shows that structure formation within virtually all parts of the unfolded protein precedes folding to the molten globule. The results

presented in this chapter, together with those reported on the molten globule of α-lactalbumin [11], show that helical molten globules apparently fold in a non-cooperative manner.

To investigate long-range interactions in unfolded apoflavodoxin that lead to formation of this off-pathway intermediate, in **chapter 4** use is made of site-directed spin labeling. For this purpose, glutamine at position 48, which resides in a non-native α-helix of unfolded apoflavodoxin, is replaced by a cysteine. This replacement enables covalent attachment of two different nitroxide spin labels, MTSL and CMTSL. Due to this amino acid replacement stability of native apoflavodoxin against unfolding decreases and attachment of the nitroxide spin label MTSL leads to a further decrease in stability. Replacement of Gln48 by Cys48 decreased flexibility of the ordered regions in unfolded apoflavodoxin in 3.4 M GuHCl, due to increased hydrophobic interactions. Interactions are detected between the MTSL spin label attached to Cys69 and region Ser40 - Leu62 of unfolded apoflavodoxin in 6.0 M GuHCl. These non-specific hydrophobic interactions between nitroxide spin labels and hydrophobic patches of unfolded apoflavodoxin perturb the unfolded protein. Our observations show that in 6.0 M GuHCl spin-labeled apoflavodoxin is less random coil than C69A apoflavodoxin is. Thus, care needs to be taken in the use of spin labels for the study of the conformational and dynamic properties of unfolded proteins. In 3.4 M GuHCl the attached CMTSL spin label induces the presence of two distinct states in unfolded apoflavodoxin. In one of these states, the spin label attached to residue 48 has persistent contact with residue Leu78. The spin label data show that non-native contacts exist between transiently ordered structured elements in unfolded apoflavodoxin.

Full population of the molten globule-like folding state of apoflavodoxin is possible through covalent introduction of just a single extra oxygen atom in the protein, achieved by replacing Phe44 with Tyr44 through site-directed mutagenesis (**chapter 5**). This replacement leads to significant destabilization of native apoflavodoxin, as is demonstrated by GuHCl-induced equilibrium (un)folding and thermal unfolding experiments. Decreasing salt concentration destabilizes native apoflavodoxin even further. As a result, the native state of F44Y apoflavodoxin is hardly populated. Instead, in absence of denaturant, virtually all protein molecules exist as molten globule-like folding intermediate. Direct characterization of this intermediate by far-UV CD is possible, it is shown that the molten globule has a totally different topology: it is helical and lacks the parallel β-sheet of native apoflavodoxin.

Full population of the molten globule state of F44Y apoflavodoxin enables use of H/D exchange for the characterization at the residue level by NMR spectroscopy of apoflavodoxin's molten globule folding intermediate. In **chapter 6**, interrupted H/D exchange is used to detect the stable core of apoflavodoxin's molten globule in absence of denaturant. Exchange rates could be determined for 68 backbone amides. Amide protons of residues Lys16 – Phe25 are poorly protected against exchange, and structure formed in this region is very unstable. In chapter 4 chemical shift data and C_m-values showed that these residues belong to the most unstable part of apoflavodoxin's molten globule, as they remain random coil down to a GuHCl concentration of 1.58 M. Leu110 to Val125 have the highest protection factors against H/D exchange and form the single stable core of apoflavodoxin's molten globule in absence of denaturant. The residues of this molten globule, which have the highest midpoints against unfolding by GuHCl, roughly coincide with those residues that are transiently ordered in unfolded apoflavodoxin. Only one of the four regions mentioned is significantly protected against exchange in this intermediate. This suggests that this helix is better buried in apoflavodoxin's molten globule compared to the other helices. Hydrophobic interactions of this helix with the other ordered parts of the molten globule, although loose in nature, cause

context-dependent stabilization of this helix against unfolding. The helical molten globule contains thus a single stable core. Non-native docking of helices in apoflavodoxin's molten globule prevents formation of the parallel β-sheet of native apoflavodoxin. Hence, to produce native α-β parallel protein molecules, the off-pathway species needs to unfold.

Discussion

Formation of non-native secondary and tertiary structure in unfolded protein is the answer to the question: "Why do α-β parallel proteins, like flavodoxins, form misfolded off-pathway intermediates?" The presence of non-native secondary structure elements in unfolded proteins is probably a widespread phenomenon. However, subsequent formation of folding intermediates that contain these non-native structure elements is likely but rarely reported.

In this thesis, it is proven for the first time that formation of native and non-native helices within an unfolded α-β parallel protein and subsequent non-native docking of these structured regions leads to formation of a compact helical off-pathway intermediate.

One of the helices (residues Leu110 to Val125) forms a stable core in the molten globule in absence of denaturant. Hydrophobic interactions of this helix with the other ordered parts of the molten globule cause its context-dependent stabilization. Non-native docking of the helices prevents formation of the parallel β-sheet of native protein. To produce native α-β parallel protein molecules, the off-pathway species needs to unfold and as a result non-native interactions and non-native secondary structure are disrupted.

This thesis shows that acquisition of native-like topology is not necessarily the general result of the initial collapse in protein folding. Rather than directing productive folding, conformational pre-organization in the unfolded state of an α-β parallel type protein promotes off-pathway species formation. The data presented in this thesis indicate that especially proteins that contain domains with an α-β parallel topology seem susceptible to off-pathway intermediate formation.

A single polypeptide sequence can code for monomeric protein folds that are largely different under native-like conditions. The amino acid sequence of apoflavodoxin codes for the α-β parallel topology of the native state, as well as for a helical protein species. Upon a mild change of conditions, topological switching between both folds occurs and a monomeric protein species with a distinct fold becomes energetically most favorable. Topological switching between unrelated protein structures is likely a general phenomenon in the protein structure universe.

References

1. Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci U S A* **103**, 4095-4100.

2. Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem* **280**, 7836-7844.

3. Bollen, Y. J., Sanchez, I. E. & van Mierlo, C. P. M. (2004). Formation of on- and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry* **43**, 10475-10489.

4. Steensma, E., Nijman, M. J., Bollen, Y. J., de Jager, P. A., van den Berg, W. A., van Dongen, W. M. & van Mierlo, C. P. M. (1998). Apparent local stability of the secondary structure of *Azotobacter vinelandii* holoflavodoxin II as probed by hydrogen exchange: implications for redox potential regulation and flavodoxin folding. *Protein Sci* **7**, 306-317.

5. Steensma, E. & van Mierlo, C. P. M. (1998). Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J Mol Biol* **282**, 653-666.

6. van Mierlo, C. P. M. & Steensma, E. (2000). Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *J Biotechnol* **79**, 281-298.

7. van Mierlo, C. P. M., van den Oever, J. M. & Steensma, E. (2000). Apoflavodoxin (un)folding followed at the residue level by NMR. *Protein Sci* **9**, 145-157.

8. van Mierlo, C. P. M., van Dongen, W. M., Vergeldt, F., van Berkel, W. J. & Steensma, E. (1998). The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Sci* **7**, 2331-2344.

9. Bollen, Y. J. & van Mierlo, C. P. M. (2005). Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181-189.

10. Rose, G. D. & Roy, S. (1980). Hydrophobic Basis of Packing in Globular-Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4643-4647.

11. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol* **4**, 630-634.

8

Nederlandse Samenvatting

Inleiding

In een menselijke cel zijn duizenden verschillende eiwitmoleculen te vinden. Daarnaast bevat een cel veel andere moleculen, zoals water en zouten, DNA, ons erfelijk materiaal, en vetzuren die celmembranen vormen. Eiwitten vormen de basis van de biologie en hebben veel verschillende functies. Het zijn bouwstenen voor structuur in bot, spieren, haren, huid en bloedvaten. Als enzymen dragen ze zorg voor biochemische reacties in onze cellen. Als antilichamen herkennen en helpen zij bij het verwijderen van vreemde indringers zoals bacteriën en virussen. In de cel zijn eiwitten aanwezig die andere eiwitten transporteren, eiwitten die functioneren als poortwachters in de celwand, eiwitten die genen aan en uit kunnen zetten en er zijn nog vele eiwitten aanwezig waarvan we de functie (nog) niet weten. De overeenkomst tussen al deze verschillende eiwitten is hun opbouw. Elk eiwit is een keten van aminozuren, de lengte van de keten varieert tussen 100 en meer dan 1000 aminozuren, echter er zijn maar 20 verschillende aminozuren waaruit de keten is opgebouwd.

Om alle verschillende taken uit te voeren moeten eiwitten gevouwen worden tot complexe drie-dimensionale structuren. Dit proces heet eiwitvouwing (Figuur 8.1), kleine eiwitten kunnen dit zelf, zonder hulp van andere eiwitten. Maar hoe weet een eiwit hoe het zich moet opvouwen en hoe gebeurt dit zonder dat er fouten worden gemaakt? Een eiwit van ongeveer 100 aminozuren is meestal binnen een seconde gevouwen tot de goede structuur, terwijl dit eiwit maar liefst 10^{89} (een 1 met 89 nullen!) verschillende vormen kan aannemen. Als het vouwen van het eiwit willekeurig zou gaan, dan duurt dit langer dan het bestaan van het universum. Deze tegenstelling heet de "paradox van Levinthal", genoemd naar de wetenschapper Cyrus Levinthal die in 1968 al bedacht heeft dat het proces van het vouwen van een eiwit niet willekeurig kan zijn, maar dat elk eiwit een "vouwingsweg" heeft en dat eiwitten ook tussenvormen (intermediairen), kunnen hebben.
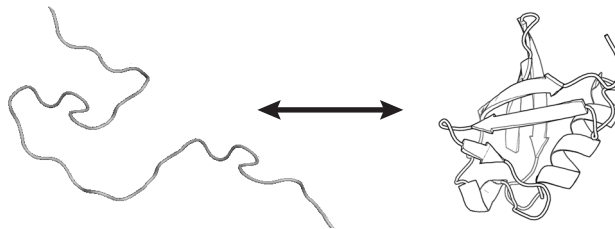

Figuur 8.1 Het vouwen van een eiwit. Links het "ontvouwen" eiwit en rechts het eiwit gevouwen in de unieke drie-dimensionale structuur van dit eiwit.

Soms gaat het vouwen van eiwitten in de cel mis. De ziekte van Alzheimer, taaislijmziekte, "gekke koeien ziekte", de ziekte van Huntington en sommige vormen van kanker worden veroorzaakt doordat eiwitten zich niet goed vouwen. Als eiwitten niet goed vouwen kan het gebeuren dat ze hun werk niet meer goed kunnen doen en dat je daardoor ziek wordt, zoals bij taaislijmziekte. Het kan ook gebeuren dat eiwitten die niet goed gevouwen zijn samen gaan klonteren en "aggregaten" vormen, waardoor je ziek wordt. Deze aggregaten worden vaak aangetroffen in de hersenen, zoals bij de ziekte van Alzheimer en BSE.

NMR spectroscopie

Een veelgebruikte en nauwkeurige techniek om eiwitten en eiwitvouwing te bestuderen is NMR (Nuclear Magnetic Resonance, oftewel Kern Spin Resonantie) spectroscopie. Met deze techniek kun je de eigenschappen van individuele atomen in een eiwit bestuderen. De atoomkernen draaien als een tol rond hun as, en gedragen zich als een klein magneetje, de kernspin. De eiwitoplossing wordt in een dun buisje in een sterk magnetisch veld gezet, het apparaat in Figuur 8.2. Daardoor zullen alle kernspins in de richting van het magnetische veld gaan wijzen. Door bestraling met hoogfrequente radiogolven kun je de richting van de kernspins beïnvloeden, dit heet het aanslaan van de kernspins. Als de bestraling stopt vallen de kernspins weer terug in hun oorspronkelijke positie waardoor ze zelf ook radiogolven uitstralen. Het uitgezonden signaal bevat informatie over de atomen die je met het experiment hebt aangeslagen. Dit zelfde principe wordt ook gebruikt bij het maken van MRI scans.
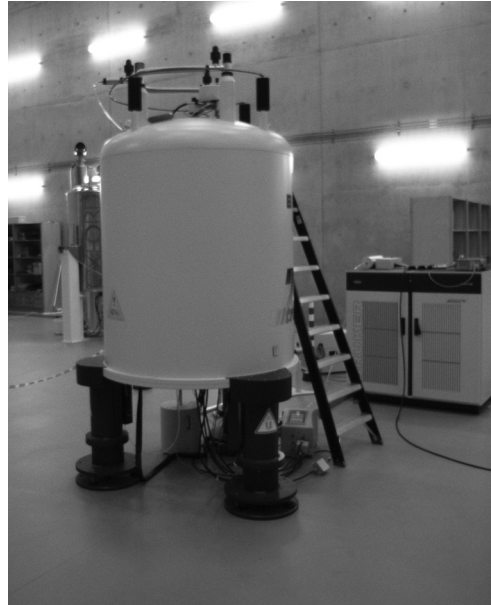
Figuur 8.2 Een NMR spectrometer. De grootte van het apparaat wordt bepaald door de magneet die er in zit. Hoe sterker de magneet hoe nauwkeuriger de eigenschappen van eiwitten gemeten kunnen worden.

Met NMR spectroscopie kunnen we verschillende eigenschappen van eiwitten meten. Een van de meest gebruikte toepassingen is het bepalen van de drie-dimensionale structuur van eiwitten. Dit gebeurt door afstanden te meten tussen verschillende waterstofatomen in het eiwit. Als je genoeg afstanden hebt, kan de computer een unieke drie-dimensionale structuur zoals in Figuur 8.3 berekenen. Voor het bestuderen van het proces van eiwitvouwing is dit geen geschikte techniek. Een ontvouwen eiwit is namelijk zo ontzettend beweeglijk (vergelijk het met roeren in een bord spaghetti) dat er geen gefixeerde afstanden tussen waterstofatomen te meten zijn. Toch kun je informatie krijgen over ontvouwen eiwitten en eiwitvouwing met NMR spectroscopie. Met NMR spectroscopie kun je namelijk heel nauwkeurig meten of er iets verandert in de omgeving van de verschillende atomen, precies wat er gebeurt als een eiwit zich vouwt! Een andere belangrijke informatiebron die verkregen wordt met NMR spectroscopie is de beweeglijkheid van de ketting van aminozuren die een eiwit vormt. De ketting van aminozuren is heel beweeglijk in een ontvouwen eiwit en veel meer gefixeerd in een gevouwen eiwit. Met deze techniek kun je bijvoorbeeld heel precies kijken in welke volgorde stukjes eiwit vouwen. Een derde toepassing van NMR spectroscopie is een methode die de uitwisseling van waterstofatomen met deuteriumatomen gebruikt. De spins van waterstofatomen kun je meten met NMR spectroscopie, de spins van deuteriumatomen (een deuteriumatoom is een waterstofisotoop dat in de kern behalve een proton ook een neutron bevat) kun je niet meten met NMR spectroscopie. De waterstofatomen in het eiwit worden alleen verwisseld als er stukjes van het eiwit ontvouwen en in aanraking komen met water. Hiermee kan bepaald worden welk stukje eiwit het beste opgevouwen en beschermd is.

Flavodoxine

In dit proefschrift probeer ik te ontdekken waarom sommige eiwitten zich (soms) verkeerd vouwen. De titel van mijn proefschrift is vrij vertaald: "Waarom vormen α-β parallelle eiwitten een misgevouwen molten globule?" Om deze vraag te beantwoorden heb ik gebruik gemaakt van het eiwit flavodoxine. Flavodoxine is een eiwit dat gemaakt wordt door de bacterie *Azotobacter vinelandii*, en het heeft als functie dat het elektronen transporteert bij biochemische reacties. Flavodoxine is gevouwen in de zogenaamde α-β parallelle topologie (Figuur 8.3). In het centrum van het eiwit bevindt zich een vlakke plaat (een zogenaamde β-sheet) met daaromheen een aantal wenteltrap-achtige constructies (de zogenaamde α-helices). Deze specifieke manier van vouwen is veel voorkomend in de natuur, eiwitten met verschillende functies hebben deze topologie. Het bijzondere aan dit eiwit en aan de andere eiwitten met deze topologie is dat het eiwit verkeerd kan vouwen. In de verkeerd gevouwen vorm van flavodoxine is het eiwit minder compact en strak opgevouwen en daardoor kunnen de eiwitten aan elkaar plakken en aggregaten vormen. Een dergelijke toestand wordt een "molten globule" toestand genoemd.

De vouwingsweg van flavodoxine is ingewikkelder dan dat van andere eiwitten. De molten globule die gevormd wordt ligt namelijk niet op de route van de ontvouwen toestand naar de gevouwen toestand, maar ligt van de weg af (off-pathway), dit betekent dat de molten globule eerst weer moet ontvouwen voordat het eiwit goed vouwt naar de natieve structuur. Schematisch ziet dit er zo uit: *Goed gevouwen* ⇆ *Ontvouwen* ⇆ *Molten Globule* Deze vouwingsweg is karakteristiek voor de α-β parallelle eiwitten en flavodoxine is daardoor een goed eiwit om deze bijzondere vouwing te bestuderen.

Figuur 8.3 Een model van het eiwit flavodoxine. De β-sheet is met pijlen weegegeven en de α-helices met de wenteltrapachtige vorm.

Samenvatting van mijn onderzoek

Ik heb onderzocht waarom flavodoxine zich verkeerd kan vouwen en hoe deze verkeerd gevouwen molten globule toestand eruit ziet. In **Hoofdstuk 2** heb ik de ontvouwen vorm van flavodoxine bestudeerd, het ontvouwen van het eiwit doe ik door een chemische denaturant (een "ontvouwingsstof") aan het eiwit toe te voegen. Een deel van het antwoord op de vraag waarom het eiwit verkeerd vouwt blijkt te vinden in de ontvouwen vorm van het eiwit. Als het eiwit ontvouwen is, blijken, als je de juiste NMR experimenten doet, er toch al α-helices in te zitten. Deze helices zijn niet dezelfde helices die voorkomen in de goed gevouwen, natieve vorm van het eiwit, maar zijn helices die voorkomen in de verkeerd gevouwen, molten globule vorm. Op de plekken waar de α-helices zich bevinden is ook een verminderde beweeglijkheid waargenomen, waarschijnlijk doordat de helices graag bij elkaar gaan zitten en daardoor minder bewegen dan de rest van het ontvouwen eiwit. In de structuur van het natieve eiwit zitten deze α-helices niet dicht bij elkaar. De eerste vouwingsstappen in het ontvouwen eiwit zijn dus de oorzaak voor de vorming van de molten globule.

In **Hoofdstuk 3** bekijk ik wat er gebeurt als het ontvouwen eiwit gaat vouwen tot de molten globule toestand door de hoeveelheid toegevoegde chemische denaturant in stappen te verlagen. We wisten al dat het vouwen van de natieve vorm van het eiwit in één stap gaat, alle stukjes eiwit vallen tegelijk als puzzelstukjes op hun plek om de natieve structuur te vormen. Het vouwen van de molten globule blijkt niet in één stap te gaan, maar in 5 verschillende stappen. Het gestructureerde deel van de molten globule wordt steeds een beetje uitgebreid als je de hoeveelheid van de chemische denaturant verlaagd. Deze vouwing in stapjes is karakteristiek voor de vouwing van molten globule eiwitten die alleen α-helices bevatten.

De eigenschappen van de kernspins in waterstofatomen in het eiwit kan beïnvloed worden door een zogenaamd paramagnetisch spinlabel. Dit is een molecuul dat op een specifieke plek in het eiwit vastgezet kan worden en het NMR signaal van alle waterstofatomen die in de buurt van het label zitten verlaagt. Door dit label aan één van de α-helices te hangen heb ik in **Hoofdstuk 4** kunnen bepalen dat hydrofobe (waterafstotende) stukken van het eiwit graag bij elkaar zitten, terwijl het eiwit is ontvouwen met een chemische denaturant. De interacties die ik gemeten heb met dit spinlabel zijn niet aanwezig in het natief gevouwen eiwit, een verdere aanwijzing voor de vorming van de molten globule die deze interacties waarschijnlijk wel heeft.

Door het muteren van het eiwit heb ik in **Hoofdstuk 5 en 6** de molten globule toestand kunnen karakteriseren. De molten globule van flavodoxine blijkt een heel andere topologie te hebben dan de gevouwen vorm van flavodoxine, het bestaat namelijk alleen uit α-helices zonder dat in het centrum van het eiwit een parallelle β-sheet wordt gevormd. De analyse van de NMR meting waarbij gekeken wordt naar de uitwisseling van waterstofatomen met deuteriumatomen heeft aangetoond dat één van de α-helices in de molten globule ingepakt zit in het midden van het eiwit, terwijl de andere α-helices meer aan de buitenkant zitten. Dit is een extra aanwijzing dat er geen centrale parallelle β-sheet wordt gevormd. Het is waarschijnlijk dat alle eiwitten die in de correct gevouwen toestand een α-β parallelle topologie hebben, een molten globule kunnen vormen met een vreemde helix-achtige topologie.

List of publications

Bollen, Y. J., Nabuurs, S. M., van Berkel, W. J. & van Mierlo, C. P. M. (2005). Last in, first out: The role of cofactor binding in flavodoxin folding. *J Biol Chem*, **280**, 7836-7844.

Engel, R., Westphal, A. H., Huberts, D. H. E. W., Nabuurs, S. M., Lindhoud, S., Visser, A. J. W. G. & van Mierlo, C. P. M. (2008). Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem*, **283**, 27383-27383.

Nabuurs, S. M., Westphal, A. & van Mierlo, C. P. M. (2008). Extensive formation of off-pathway species during folding of an α-β parallel protein is due to docking of (non)native structure elements in unfolded molecules. *J Am Chem Soc*, **130**, 16914-16920.

Nabuurs, S. M., Westphal, A. & van Mierlo, C. P. M. (2008). Non-cooperative formation of the off-pathway molten globule during folding of the α–β parallel protein apoflavodoxin. *J Am Chem Soc* **131**, 2739-2746.

Bedankt!

Nu mijn proefschrift bijna af is en bijna naar de drukker gaat, heb ik tijd om iedereen te bedanken die mij de afgelopen 5 (!) jaar op de een of andere manier heeft geholpen om deze promotie tot een goed einde te brengen.

Allereerst wil ik Carlo van Mierlo mijn co-promotor en dagelijks begeleider bedanken. Carlo, zonder jouw inspirerende manier van begeleiden was dit proefschrift niet zo mooi geworden. In het eerste jaar heb ik moeten wennen aan manier van denken, maar al snel zaten we op dezelfde (wetenschappelijke) golflengte. Ik vind het bijzonder fijn dat je me zoveel vrijheid in mijn onderzoek hebt gegeven en dat ik altijd binnen kon lopen met vragen en opmerkingen over wat dan ook. We hebben heel wat uren pratend doorgebracht in jouw overvolle kamertje, gelukkig ging het niet altijd over wetenschap alleen! Op de momenten dat ik het even niet zag zitten met mijn onderzoek waren er altijd weer motiverende en relativerende woorden waar ik weer nieuwe energie uit haalde. Op de momenten dat we een doorbraak hadden of ons werk door anderen gewaardeerd werd maakte jouw enthousiasme mij extra blij. Ik heb veel van je geleerd, concrete vaardigheden zoals het schrijven van artikelen en het geven van presentaties, maar ook hoe je heftige discussies kan hebben over wetenschap en daarna samen koffie gaat drinken. Bedankt voor dit alles!

Mijn promotor Sacco de Vries wil ik bedanken voor de mogelijkheid om bij het laboratorium voor biochemie te promoveren. Jouw 'airplane-view' op mijn onderzoek was vaak relativerend en ik zal de gezellige gesprekken bij de koffie en andere sociale activiteiten niet vergeten.

Zonder Adrie geen eiwit en dus geen proefschrift... Ik heb heel veel van je geleerd (en zou nog veel meer kunnen leren) in en buiten het lab. Je bent ook een ontzettend gezellige collega (altijd een goede grap paraat) en altijd bereid om met alles te helpen. Ik hoop dat je nog lang met zoveel plezier bij biochemie blijft werken en soms aan mij denkt als je de schroefjes van de Åkta losdraait...

Simon wil ik bedanken voor de snoepjes (de 'gewone' zijn het lekkerst!). En omdat je altijd mee wilde koffie drinken, omdat je een van de leukste studenten was die ik met AMBR heb gehad, omdat je het laatste jaar al mijn "laatste jaars" beslommeringen (stress, solliciteren, deadlines, nog meer stress, nog meer solliciteren) zonder morren hebt aangehoord en omdat ik zoveel met (en om?) je heb moeten lachen. En natuurlijk omdat je mijn paranimf wilt zijn en voor mij een apenpakje aantrekt! Over een paar jaar zal ik ook trots in de aula zitten om jouw promotie mee te maken!

Ruchira, we shared a room for more than three years. Already after one month I attended your wedding and a year later your PhD defence. Every afternoon we used our little electric kettle to make hot water for thee (for me) and instant coffee (with lots of cream and sugar!) for you. I really enjoyed the discussion about flavodoxin folding we had when you joined our group as a post-doc, but even more I enjoyed our daily small talk. Then you became pregnant from Leon and went on pregnancy leave. Luckily Ede wasn't far by car, so I could visit you from time to time. When you left The Netherlands with Maarten and Leon to work in Leeds, it became very quiet in my room. I still miss the vivid conservations we had, I'm glad I made a friend!

Er zijn veel mensen op het lab die ik moet bedanken. Yves, je was eigenlijk al vertrokken toen ik kwam, maar ik weet nog dat ik erg onder de indruk was toen ik de leesversie van jouw proefschrift zag (kan ik dat ook???). Jouw ontdekking van de "off-pathway intermediate" is de basis waarop mijn proefschrift staat. Ik wil je bedanken voor alle goede gesprekken die we over eiwitvouwing hebben gehad (en alle gezellige gesprekken zoals in Londen!) en voor alle manuscripten die je hebt gelezen! Jan Willem, bedankt voor alle gezelligheid, jij kan mensen aan het lachen maken zelfs als ze er geen zin in hebben. Ik wil je ook bedanken voor je gastvrijheid (als het stormt...) en voor je danskunsten. Op iedere werkplek heb je eigenlijk een Jan Willem nodig... Ton, bedankt voor alles wat ik van je heb geleerd over fluorescentie en de wetenschap. Door de werkbesprekingen met jou probeer ik ook door een andere bril naar mijn eigen werk te kijken, het was erg inspirerend om met je samen te werken. Nina, leuk dat jij bij ons in de eiwitvouwingsgroep kwam werken. Willy, bedankt voor alle hulp die je me hebt gegeven bij de Åkta als het weer eens fout ging en voor alle praktische tips op het lab. Willem, jij keek altijd van een afstand naar ons onderzoek, dat resulteerde in goede gesprekken over onderzoek en vooral in gezellige gesprekken op feestjes en borrels. Sjef, we hebben niet veel samen gewerkt (behalve thee drinken en lunchen), maar wel een MALDI-TOF meting zonder te praten (keelontsteking?). Laura, zonder goede secretaresse is een afdeling niks (weet ik nu uit eigen ervaring), bedankt voor alle hulp met alles wat ik niet wist of niet kon. Huub, jij was een goede "buurman". Ik heb veel van je geleerd tijdens mijn tijd in de OR, maar ook bij de koffie was het altijd gezellig.

Vijf jaar geleden kwam ik vanuit Utrecht naar Wageningen, ik wil een paar mensen uit die eerste tijd noemen: Marielle, jouw aanwezigheid was een constante factor, koffie, lunch en thee precies op tijd. Maar het was altijd gezellig met jou! Niels, we begonnen samen met eiwitvouwing, op twee verschillende verdiepingen, maar het bleek niet jouw promotieonderzoek te zijn, nog veel succes in Amsterdam! Yee, het was vaak onduidelijk of je wel op het lab was, maar lunchen deed je altijd met ons! Monique, we zijn tegelijk begonnen en deelden een lab, alleen niet voor lang. Jose, jij bent nog niet zolang weg, ik denk aan je als een appeltje eet bij de koffie! Anja, jij was een gezellige kamergenoot, al onze dagelijkse beslommeringen en leuke weekenden werden besproken, succes met je nieuwe baan. Nog meer oud-collega's aan wie ik goede herinneringen heb: Mark Kwaaitaal, Mark Hink, Casper Vroemen (ook Utrecht...), Vief (hoe is je nieuwe baan?), Isabella, Marloes en Kees-Jan.

Er werken gelukkig nog steeds leuke mensen in het lab; Nicole, het was leuk om de laatste loodjes te kunnen delen, heel veel succes op 8 april, ik ga ontspannen in de aula zitten... Rumyana, your defence was my rehearsal, I hope mine will be as good as yours! Eike, I will remember the WK voetbal we went to! Stefania, thank you for the tiramisu, ciao! Sofia, it was nice to have someone around who knew what the abbreviation NMR meant. Justin, Wilma, Barbara, Annemarie, Dolf, Cathy, Walter, Jacques, Boudewijn, Hans en iedereen die ik nu vergeet, allemaal bedankt voor de leuke tijd die ik heb gehad bij biochemie!

Ik wil ook de studenten die met mij gewerkt hebben bedanken. Ik was nog geen jaar bezig en ik kreeg een studente Moleculaire Wetenschappen "onder mijn hoede". Daphne, ik vond het erg leuk om jou te begeleiden, hoewel, in het begin werden wij vooral samen begeleid door Adrie! Je hebt met je werk de basis gelegd voor de fluorescente flavodoxine moleculen. Toen Bregje bij mij een afstudeervak kwam doen was ik erg blij dat er een student geïnteresseerd was in NMR spectroscopie,

de meetdagen in Utrecht waren dan ook erg gezellig. Je hebt in vier maanden veel werk verricht en daarmee de basis gelegd voor een van mijn hoofdstukken. De flavodoxine variant waarmee we de molten globule hebben kunnen karakteriseren heb ik te danken aan Marije, de weken dat jij achter een computerscherm naar het eiwit hebt gekeken zijn zeker niet voor niks geweest! Nanda je was niet mijn student, maar het was echt leuk dat je er was en dat je meegeholpen hebt aan ons eiwitvouwingsonderzoek.

Buiten het lab zijn er ook mensen die ik wil bedanken. Eén verdieping lager bij biofysica zitten mannen waarvan ik hulp heb gekregen: Arie van Hoek, de tijdsopgeloste fluorescentie metingen die ik in de kelder heb gedaan waren de meest relaxte experimenten ooit (dankzij de muziek?). Frank Vergeldt, alle problemen met de Unix systemen wist jij op te lossen. Pieter de Waard, ik heb toch nog even met het nieuwe NMR systeem mogen spelen.

Ik heb veel NMR experimenten in Utrecht bij de biomoleculaire NMR groep van Rolf Boelens gedaan. Rolf, bedankt voor alle meettijd en vooral bedankt voor alle hulp (en geduld) in het begin met het opzetten van de experimenten (het ging nog wel eens fout). Ik ben in aanraking gekomen met de NMR spectroscopie door mijn afstudeervak bij Klaartje, jij hebt me veel geleerd over NMR spectroscopie wat vooral het begin van mijn promotieonderzoek een stuk makkelijker heeft gemaakt. De eerste experimenten heb ik samen met Hans Wienk opgezet, maar ook Tammo, Hugo en natuurlijk Rainer hebben me geholpen met de experimenten, zonder jullie was het niet gelukt! Gert bedankt dat ik op je lab mocht rondscharrelen en Barbara bedankt dat je voor mij zo vaak een pasje hebt geregeld. Jenny, ons tripje naar Kopenhagen was gezellig, ik heb volgens mij nog nooit met iemand in zo'n kleine hotelkamer geslapen! Tsjerk, we hebben een poging gewaagd om de vouwing van flavodoxine te simuleren.

Ik heb ook nog twee keer een weekje in Groningen rondgelopen. Frans, ik vond het erg leuk om met je samen te werken, heb ik toch ook een keer op een Varian gemeten! Wie weet lukt het in de toekomst nog om iets interessants aan flavodoxine te meten. Renee, leuk dat ik jou heb leren kennen, ik kom zeker nog een keer naar Groningen om een biertje te drinken!

Arie, Hans, Margaret en de rest van de OR van AFSG, bedankt voor de leuke en leerzame tijd die heb gehad in de OR.

Gelukkig is promoveren niet alleen maar werken. Ik wil ook graag de mensen bedanken waarmee ik in mijn vrije tijd leuke dingen heb gedaan! Mijn huisgenootjes van Warande 144 en 191, onze flatgenootjes, Sjef en Judith, Steven en Jet (als ik doctor ben, win ik dan wel met Risk?). Jannes, Mara, Raoul, Chantal, Liesbeth en Sipke, de etentjes zijn altijd gezellig (we hebben bijna een 10-jarig jubileum!). Hanneke, jouw (promotie)reis begint net, wanneer gaan we weer naar een musical? Roel, we moeten snel bij kletsen! Angelique (grappig wat onze dromen waren en waar we nu gekomen zijn), Jeroen, Sofie en ???! Suzanne en Jasper, bijna met z'n drietjes, we gaan lekker genieten van de tuin.

Waar ben je zonder familie... Foppe en Ida, bedankt voor jullie belangstelling en medeleven met mijn promotie, in jullie huis in het mooie Friesland heb ik meer dan eens de rust kunnen vinden om te schrijven. Jorrit (kijk je via internet?) en Fardau (zusje!), het is altijd gezellig (druk) als we met z'n allen zijn!

Remco bedankt dat je mijn paranimf wilt zijn, toen we nog klein waren was jij degene die de interesse in de wetenschap bij me aanwakkerde. Iris, bij jullie in Amsterdam is het altijd heerlijk ontspannen, misschien zijn het de katten?

Papa en Mama, ik heb niet genoeg woorden om jullie te bedanken. Ik heb zoveel van jullie meegekregen. Eén van de lessen: 'Ik kan het!' is mij altijd bijgebleven en heeft me vaak geholpen om door te zetten. Bedankt dat jullie nooit aan me twijfelen en me overal in steunen!

Giant steps are what you take... Bouwe Frank, bedankt voor jouw liefde en steun. Elke ochtend naast jou wakker worden is het mooiste wat me ooit is overkomen.

   Sanne

Curriculum Vitae

Sanne Nabuurs werd geboren op 24 april 1979 in Venray en groeide op in Overloon. In 1997 haalde zij haar Gymnasium diploma aan het Raayland College te Venray. In september 1997 startte zij met de studie Scheikunde aan de Universiteit Utrecht. Tijdens haar studie heeft zij een afstudeervak gedaan in de groep van Prof. dr. Rolf Boelens, Biomoleculaire NMR spectroscopie aan de Universiteit Utrecht. Dit onderzoek was gericht op de ontwikkeling van een nieuwe NMR methode om de dynamica van zijgroepen in eiwitten te bestuderen. Daarnaast heeft zij in de groep van Prof. Boelens een literatuurscriptie geschreven over de validatie van eiwitstructuren opgelost met NMR spectroscopie. In april 2004 haalde zij haar doctoraaldiploma, waarna zij begon als promovendus aan Wageningen Universiteit in de groep van dr. ir. Carlo van Mierlo in het laboratorium voor Biochemie, geleid door Prof. dr. Sacco de Vries. Dit promotieonderzoek was gericht op de vouwing en misvouwing van het eiwit flavodoxine met behulp van NMR spectroscopie en heeft geleid tot dit proefschrift.

Overview of completed training activities

Discipline specific activities
Courses

International Advanced Course Bio-nanotechnology, VLAG	2005
Advanced NMR spectroscopy, Bijvoet-NSR Research Aliance	2005
Protein misfolding Summer school Spetsai, EMBO/FEBS	2005
Capita Selecta "Structure and dynamics of biomolecules in vitro and in vivo",	
Lab of Biophysics, WUR	2006
Bruker NMR workshop, UU	2006

Meetings

From Molecule to Cell, Kick-off meeting, NWO	2004
European Experimental NMR conference	2005
Protein/Biophysics Meetings NWO, NWO	2004-2007
+ Lectures (2007)	
Annual Meeting of the Dutch NMR-DG, NMR-DG	2006-2007
+ Lecture (2007)	
Scientific symposium in honour of Prof. Kaptein, UU	2006
6th European Biophysics Congress, European Biophysics Societies' Association	2007
Belgian Biophysical Society Symposium, Belgian Biophysical Society	2007
Chaperone meeting, UU	2007
+ Lecture	
Chaperone meeting, VU	2008
NMR-Life 2007, University of Copenhagen	2007

General courses

VLAG PhD week	2004
NWO Talent Day	2005
Presentation Course, Wageningen University	2004
R&O training, WUR HRM	2004